

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:56:07 ; Search time 53 Seconds
(without alignments)
4306.581 Million cell updates/sec

Title: US-10-047-257-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSWDYMQSD.....VWQIALRMEVLGCEAQLDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_13Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7691	100.0	1438	21	AA01262
2	7691	100.0	1457	20	AAV21675
3	7684	99.9	1457	19	AAW46246
4	7684	99.9	1457	19	AAW43372
5	7674	99.8	1471	18	AAW23414
6	7674	99.8	1471	22	AAW67959
7	7659.5	99.6	1440	12	AAE12971
8	7626	99.2	1459	22	AAE10832
9	7616	99.0	1459	22	AAE10833

10	7613	99.0	1459	22	AAE10827	Human factor VIII
11	7611.5	99.0	1516	9	AA080265	Modified factor VI
12	7592	98.7	1424	22	AA048842	Mutant mature huma
13	7592	98.7	1424	23	AA018622	Human mature B-dom
14	7588	98.7	1447	23	ABG92541	Sarg B-domain-dele
15	7581	98.6	1425	9	AA080267	Modified factor VI
16	7578	98.5	1424	9	AA080268	Modified factor VI
17	7578	98.5	1445	23	ABG92540	LE B-domain-dele
18	7574	98.5	1424	10	AA091169	Sequence of 740 Ar
19	7434	96.7	1661	18	AAW18670	Factor VIII-dB695-
20	7353.5	95.6	1383	18	AAW33227	Procoagulant-activ
21	7345.5	95.5	1383	18	AAW33229	Procoagulant-activ
22	7342.5	95.5	1383	18	AAW33228	Procoagulant-activ
23	7234	94.1	2332	14	AAW43257	Human Factor VIII.
24	7234	94.1	2332	19	AAW3483	Human factor VIII.
25	7234	94.1	2332	20	AAW31594	Human factor VIII
26	7234	94.1	2332	22	AAE10826	Human mature wild-
27	7234	94.1	2332	22	AAE11200	Human factor VIII
28	7234	94.1	2332	22	AAW50465	Human factor VIII.
29	7234	94.1	2332	23	AAU79869	Human factor VIII.
30	7234	94.1	2351	20	AAW21676	Factor VIII protei
31	7231.5	94.0	2342	18	AAW11422	Active Factor VIII
32	7231	94.0	2351	18	AAW10591	Factor VIII:C (Phe
33	7231	94.0	2351	18	AAW13496	Factor VIII:C (Arg
34	7230.5	94.0	2344	18	AAW11432	Active Factor VIII
35	7230.5	94.0	2344	18	AAW11410	Active Factor VIII
36	7230	94.0	2332	23	AAU79870	Human factor VIII
37	7230	94.0	2332	23	AAU79872	Human factor VIII
38	7230	94.0	2345	18	AAW11415	Active Factor VIII
39	7230	94.0	2345	18	AAW11403	Active Factor VIII
40	7229.5	94.0	2346	18	AAW11421	Active Factor VIII
41	7229.5	94.0	2346	18	AAW11431	Active Factor VIII
42	7229.5	94.0	2346	18	AAW11434	Active Factor VIII
43	7229	94.0	2347	18	AAW11411	Active Factor VIII
44	7229	94.0	2347	18	AAW11402	Active Factor VIII
45	7229	94.0	2351	18	AAW10592	Factor VIII:C (Tyr

ALIGNMENTS

RESULT 1

AA01262

ID AAB01262 standard; protein; 1438 AA.

XX AAB01262;

XX AAB01262;

DT 25-SEP-2000 (first entry)

XX B-domain deleted factor VIII sequence.

DE Factor VIII; procoagulant; adenovirus; adeno-associated strain;
KW Gene therapy; human Burkitt's lymphoma; HKD; therapy;
KW Therapeutic protein; vector; Epstein-Barr virus; human.

XX Homo sapiens.

XX WO200034505-A1.

XX 15-JUN-2000.

PF 08-DEC-1999; 99WO-US29169.

PR 10-DEC-1998; 98US-0209916.

PA (FARB) BAYER CORP.

PI Cho M, Chan SY, Kelsey W, Yee H;

DR WPI; 2000-431311/37.

XX Producing cells expressing a protein having factor VIII procoagulant
PT activity especially, human factor VIII in an industrial scale, involves

PT expressing a vector comprising a sequence coding for factor VIII in
 XX human cells

Claim 7; Fig 1; 27pp; English.

CC Producing cells expressing a protein having factor VIII procoagulant
 CC activity, comprises contacting the cells with a vector comprising a
 CC selectable marker and a sequence coding for the protein having factor
 CC VIII procoagulant activity operably linked to a promoter. The cells
 CC are then selected and individual clones expressing high levels of the
 CC protein are isolated from the selected cells. The cells produced by
 CC the method are not only useful for producing protein having factor
 CC VIII procoagulant activity but also for producing adenovirus and
 CC adeno-associated virus strains for gene therapy. The advantage of
 CC having cells producing protein with factor VIII procoagulant activity
 CC is that factor VIII protein can be produced on an industrial scale
 CC in the range of 2-4 pg/cell/day. Human Burkitt's lymphoma (HKB)
 CC cells provide a protein-free production system to produce not only
 CC B-domain deleted factor VIII but also other therapeutic proteins. The
 CC vector used in the method preferably comprises B-domain deleted
 CC factor VIII (BDD-FVIII), a transcriptional unit for BDD-FVIII and a
 CC selectable marker, dihydrofolate reductase (dhfr). In addition, a
 CC terminal repeat sequence from Epstein-Barr virus is inserted into the
 CC vector to increase integration efficiency.

XX Sequence 1438 AA;

Query Match 100.0%; Score 7691; DB 21; Length 1438;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATRRYVGLGAVELSDYMQSDLGELPVDARFPVRPKSPFPNTSVYKKTFLVEFTVHLEN 60
 DB 1 ATRRYVGLGAVELSDYMQSDLGELPVDARFPVRPKSPFPNTSVYKKTFLVEFTVHLEN 60
 QY 61 IAKPRPPMGLGPTIAQVYDTVTITLKNMASHPVSLHVGVSYWKASEGAYDDQTSQ 120
 DB 61 IAKPRPPMGLGPTIAQVYDTVTITLKNMASHPVSLHVGVSYWKASEGAYDDQTSQ 120
 QY 121 REKEDDKVPVGGSHYTVMOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 DB 121 REKEDDKVPVGGSHYTVMOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 QY 181 EGS LAKEKQTQTLKRTLLFAVDEKSWHSETKNSLMODRDAASARAPKHTVNGYNR 240
 DB 181 EGS LAKEKQTQTLKRTLLFAVDEKSWHSETKNSLMODRDAASARAPKHTVNGYNR 240
 QY 241 SLFGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 300
 DB 241 SLFGLIGCHRSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 300
 QY 301 MDLGFLLFCHTSSHQHGMAYVKVDSCEPBPQLRMKNNEAEYDDDLTSDMDVVRP 360
 DB 301 MDLGFLLFCHTSSHQHGMAYVKVDSCEPBPQLRMKNNEAEYDDDLTSDMDVVRP 360
 QY 361 DDNSPSTQIRSVAKHPKTVHVIYAAEEDWDYAPLVLPDDRYSKSOYLNNQPORIG 420
 DB 361 DDNSPSTQIRSVAKHPKTVHVIYAAEEDWDYAPLVLPDDRYSKSOYLNNQPORIG 420
 QY 421 RYKVKVRMAYTDETFKTRTALIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
 DB 421 RYKVKVRMAYTDETFKTRTALIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
 QY 481 TDVRLYRRRLPKGVKHLKOPFLPEGEIFKYKWTVTVEDGTPKSPRLCTRIYSFVNM 540
 DB 481 TDVRLYRRRLPKGVKHLKOPFLPEGEIFKYKWTVTVEDGTPKSPRLCTRIYSFVNM 540
 QY 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
 DB 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
 QY 601 VOLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
 DB 601 VOLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660

DB 601 VOLEDPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
 QY 661 KMVYEDTLTLFPFSGETVFMSENPGLWILCHNSDPRNRGTALLKYSSCDKNTGDYVE 720
 DB 661 KMVYEDTLTLFPFSGETVFMSENPGLWILCHNSDPRNRGTALLKYSSCDKNTGDYVE 720
 QY 721 DSYEDISAYLSKNAIEPRFSQNPVPLVKHQREITRTTLQSDQEEIDYDDTISVEMKK 780
 DB 721 DSYEDISAYLSKNAIEPRFSQNPVPLVKHQREITRTTLQSDQEEIDYDDTISVEMKK 780
 QY 781 EDFDIYDEDENQSPRSFQKKTRHYPIAAVERLWDYCHSSSPHVLNRAQSGSVPOFKKV 840
 DB 781 EDFDIYDEDENQSPRSFQKKTRHYPIAAVERLWDYCHSSSPHVLNRAQSGSVPOFKKV 840
 QY 841 FQEFDTGDSFTQPLRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRYSYSSLSIYEE 900
 DB 841 FQEFDTGDSFTQPLRGELNEHLGLLGPYIRAEVDENIMVTFRNOASRYSYSSLSIYEE 900
 QY 901 DQQAEPKPKVFPNETKTYFWKVQHMAPTKDEFCCKAWAYTSDVLEKDVHSGSLIGP 960
 DB 901 DQQAEPKPKVFPNETKTYFWKVQHMAPTKDEFCCKAWAYTSDVLEKDVHSGSLIGP 960
 QY 961 LLVCHTNTLNPAHGRQVTVQBFALFFTIFDETKSWYFTENMRNCRAPCNIQMEDPTPK 1020
 DB 961 LLVCHTNTLNPAHGRQVTVQBFALFFTIFDETKSWYFTENMRNCRAPCNIQMEDPTPK 1020
 QY 1021 NYRFHAINGYTMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFHSGHVTVTKKEEYKMA 1080
 DB 1021 NYRFHAINGYTMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFHSGHVTVTKKEEYKMA 1080
 QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKKCOTPLGMASGHIRD 1140
 DB 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKKCOTPLGMASGHIRD 1140
 QY 1141 FOITASQYQGWAPKLARLHVSGSINAWSTKEPESWIKVDLLAPMIHGIKTQGAROKFS 1200
 DB 1141 FOITASQYQGWAPKLARLHVSGSINAWSTKEPESWIKVDLLAPMIHGIKTQGAROKFS 1200
 QY 1201 SLYISQFIIMYSLDGKKWQTYRGNSGTGLMVFNGVDSGSGIKKNI FNPPIIARIHLPT 1260
 DB 1201 SLYISQFIIMYSLDGKKWQTYRGNSGTGLMVFNGVDSGSGIKKNI FNPPIIARIHLPT 1260
 QY 1261 HYSIRSTRMLMGLMCDLNSCMPLGMSKAI SDAQITASSYFTNMFATWSPSKARLHLQ 1320
 DB 1261 HYSIRSTRMLMGLMCDLNSCMPLGMSKAI SDAQITASSYFTNMFATWSPSKARLHLQ 1320
 QY 1321 RSNARWPOVANNPKEWLQVDFOKTMKVTTQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
 DB 1321 RSNARWPOVANNPKEWLQVDFOKTMKVTTQGVKSLTSMYKVEFLISSQDGHQWTLF 1380
 QY 1381 FQNGKVKVQGNQSDFTPVVNSLDPRLTRYLRTHPQSWHQAIALRMEVLGCEAQDLY 1438
 DB 1381 FQNGKVKVQGNQSDFTPVVNSLDPRLTRYLRTHPQSWHQAIALRMEVLGCEAQDLY 1438

RESULT 2

AA121675

ID AA121675 standard; Protein; 1457 AA.

XX AA121675;

XX 18-AUG-1999 (first entry)

DT Beta-domain deleted Factor VIII protein.

DE Factor VIII protein; gene modification; gene therapy; clinical disorder;

XX splicing pattern; RNA processing; gene regulation; beta-domain; human.

XX Homo sapiens.

XX WO9929848-A1.

XX 17-JUN-1999.

PD

XX PF 25-NOV-1998; 98WO-US25354.
XX PR 16-JAN-1998; 98US-0071596.
XX PR 05-DEC-1997; 97US-0067614.
XX PA (IMMU-) IMMUNE RESPONSE CORP.
XX Bidlingmaier S, Gonzales JEN, Ill CR, Yang CQ;
XX WPI: 1999-385602/32.
XX N-PSDB; AAX82258, AAX82259, AAX82260.
XX Genes and vectors exhibiting increased expression and novel splicing
XX PT patterns, useful for expression of, e.g. beta-domain deleted factor
XX VIII
XX Disclosure; Page 72-78; 123pp; English.
XX The invention describes novel genes and vectors exhibiting increased
XX expression and novel splicing patterns. It provides a gene encoding a
XX Factor VIII protein, that comprises one or more consensus or near
XX consensus splice sites which have been corrected to increase expression.
XX The method, DNA sequences and expression vectors can be used to increase
XX the expression of a gene, especially a Factor VIII gene. Genes containing
XX modified 5' and/or 3' untranslated regions have optimized expression
XX levels and tissue-specific expression. The methods are used for
XX identification and correction of consensus splice sites, addition of
XX introns, optimization of 5' and 3' untranslated regions and increase in
XX cytoplasmic RNA accumulation. Hence the DNAs are useful in gene therapy
XX to treat a clinical disorder, to study RNA processing and/or gene
XX regulation. The present sequence represents a beta-domain deleted Factor
XX VIII protein.
XX Sequence 1457 AA;
Query Match 100.0%; Score 7691; DB 20; Length 1457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATRRYVLGAVELSDWYQSDLGELPVDAREPRVPKSPFNTSVYKTLFVEFTVHLN 60
DB 20 ATRRYVLGAVELSDWYQSDLGELPVDAREPRVPKSPFNTSVYKTLFVEFTVHLN 79
QY 61 IAKPRPPMGLLGPITQAEVDTVTITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120
DB 80 IAKPRPPMGLLGPITQAEVDTVTITLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 139
QY 121 REKEDKVPFGSGSHYVWQVLENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDKVPFGSGSHYVWQVLENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS LAKEKTQTLLKFTLLFAVDEGKSWHSEKNSLMODRDAASARAWPKHVTNGVYNR 240
DB 200 EGS LAKEKTQTLLKFTLLFAVDEGKSWHSEKNSLMODRDAASARAWPKHVTNGVYNR 259
QY 241 SLFGLIGCHRSYVWHVIGMGTTPVHSIFLEHTFLVRNHRQASLEISPIITFLTAQTLL 300
DB 260 SLFGLIGCHRSYVWHVIGMGTTPVHSIFLEHTFLVRNHRQASLEISPIITFLTAQTLL 319
QY 301 MDLGGFLFCHLSSHQHGMAYVKVDSCEPQLRMKNNEAEYDDDLTDSMDVVRP 360
DB 320 MDLGGFLFCHLSSHQHGMAYVKVDSCEPQLRMKNNEAEYDDDLTDSMDVVRP 379
QY 361 DDNSPSFIQIRSAKHPKTVWHVYIAAEEEDMDYAPLVAPDRSRYKSYQLNNGPORG 420
DB 380 DDNSPSFIQIRSAKHPKTVWHVYIAAEEEDMDYAPLVAPDRSRYKSYQLNNGPORG 439
QY 421 RYKVKVRMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 480
DB 440 RYKVKVRMAYTDETFKTREAIQHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 499
QY 481 TDVRLYSRRLPKGVKHLKDFPILFGEIFKYKWTTVEDGPKSDPRCLTRYSSFVWME 540

DB 500 TDVRLYSRRLPKGVKHLKDFPILFGEIFKYKWTTVEDGPKSDPRCLTRYSSFVWME 559
QY 541 RDLASGLTGPLLI CYKESVDQRGNQIMSDKRNVIILFVSPDENRSWYLTENIORFLPNPAG 600
DB 560 RDLASGLTGPLLI CYKESVDQRGNQIMSDKRNVIILFVSPDENRSWYLTENIORFLPNPAG 619
QY 601 VOLEPPEQASNIMHSINGYVFDLSIQLSVCLHEVAYWYLTLSIGATDRLSVPFSGVTPGH 660
DB 620 VOLEPPEQASNIMHSINGYVFDLSIQLSVCLHEVAYWYLTLSIGATDRLSVPFSGVTPGH 679
QY 661 KMVYEDTLTLFPFSGETVFMSENFGMLITGLCHNSDFRNRGMTALLKVSCKDKNTGDIYE 720
DB 680 KMVYEDTLTLFPFSGETVFMSENFGMLITGLCHNSDFRNRGMTALLKVSCKDKNTGDIYE 739
QY 721 DSYEDISAYLLSKNNAIEPRSPQNPVILKHORITRTTLOSDEEDIDYDTISVEMKK 780
DB 740 DSYEDISAYLLSKNNAIEPRSPQNPVILKHORITRTTLOSDEEDIDYDTISVEMKK 799
QY 781 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAGSGSVPOPKV 840
DB 800 EDFDIYDEDNQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAGSGSVPOPKV 859
QY 841 FOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYEE 900
DB 860 FOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASRPYSFYSSLSIYEE 919
QY 901 DORQAEPRKPNVKNETKTYFKVQOHMAKDFDKANAFSDVDLEKDVHSGLTGP 960
DB 920 DORQAEPRKPNVKNETKTYFKVQOHMAKDFDKANAFSDVDLEKDVHSGLTGP 979
QY 961 LLVCHTNTLNPAHQGVQVQFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKE 1020
DB 980 LLVCHTNTLNPAHQGVQVQFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKE 1039
QY 1021 NYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENENIHSIFSGHVTFRKKEEYKMA 1080
DB 1040 NYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENENIHSIFSGHVTFRKKEEYKMA 1099
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCOTPLGMASGHIRD 1140
DB 1100 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCOTPLGMASGHIRD 1159
QY 1141 FOITASQGYGOWAPKARLHYSGSTINAMSTKEPFSWIKVDLLAPMIHGIKTQGAROKFS 1200
DB 1160 FOITASQGYGOWAPKARLHYSGSTINAMSTKEPFSWIKVDLLAPMIHGIKTQGAROKFS 1219
QY 1201 SLYISQFTIMYSLDGKMKQTYRGNTGTLMVFFGNVDSGSGIKHINFPPIIARIYIRLHPT 1260
DB 1220 SLYISQFTIMYSLDGKMKQTYRGNTGTLMVFFGNVDSGSGIKHINFPPIIARIYIRLHPT 1279
QY 1261 HYSIRSTLRMELMGCDLNSCSMPGLMESKATSDAQITASSYFTNMFATWSKARLHLQ 1320
DB 1280 HYSIRSTLRMELMGCDLNSCSMPGLMESKATSDAQITASSYFTNMFATWSKARLHLQ 1339
QY 1321 RSNARPOVNPKNLEQVDFQTKMKVGTVTQGVKSLTSMYVKEFLTSSSDQGHQWTLF 1380
DB 1340 RSNARPOVNPKNLEQVDFQTKMKVGTVTQGVKSLTSMYVKEFLTSSSDQGHQWTLF 1399
QY 1381 FQNGKVKVQGNQSDFTFPVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1438
DB 1400 FQNGKVKVQGNQSDFTFPVNSLDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAQDLY 1457
RESULT 3
AAW46246
ID AAW46246 standard; Protein; 1457 AA.
XX AAW46246;
XX
XX
DT 25-MAR-2003 (updated)
DT 06-AUG-1998 (first entry)
XX

Human factor VIII beta-domain deleted SQN deletion protein sequence.

Replication defective; recombinant retrovirus; RRV; therapeutic protein; haemophilia; thrombosis; hypercoagulable disorder; liver disease; human; hepatitis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes; cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia; hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia; Guacher's syndrome; high blood pressure; Alzheimer's disease, autoimmune; inflammatory disease; factor VIII.

Homo sapiens.

W09800541-A2.

08-JAN-1998.

02-JUL-1997; 97WO-US11784.

04-JUN-1997; 97US-0869309.

03-JUL-1996; 96US-0645601.

13-AUG-1996; 96US-0696381.

(CHIR) CHIRON CORP.

Jolly DJ, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M; Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE; Mittelsaetd DM, Fruessak CE, Greengard J, Lee R;

WPI; 1998-086966/08.

N-PSDB; AAV19581.

New replication defective recombinant retro-viruses - which can be administered to provide long term systemic expression of therapeutic protein in blood, useful in, e.g. treating hyper-coagulable disorders

Example 28; Pages 213-217; 272pp; English.

This is the beta-domain deleted SQN deletion protein of human factor VIII. The encoding DNA is used to construct recombinant retroviral vectors expressing human factor VIII. The invention provides the preparation of replication defective recombinant retrovirus (RRV) expressing a therapeutic protein. The RRV preparation is resistant to degradation by human complement and is capable of inducing long term systemic expression of the therapeutic protein when administered intravenously to a human. The long term systemic expression results in a measurable level of the therapeutic protein being produced in the blood of the human for a period of at least 30 days after the administration of the RRV vector preparation. RRV's can be used for in vivo delivery of therapeutic protein to treat, e.g. haemophilia A, haemophilia B, thrombosis, hypercoagulable disorders, liver diseases such as hepatitis, disorders such as thalassemia, phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency (SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited emphysema, familial hypercholesterolemia, diabetes, hypopituitarism, adenine deaminase deficiency, alphas-1 antitrypsin deficiency, Guacher's syndrome, anaemia, infections such as HIV infection, high blood pressure, Alzheimer's disease, autoimmune or inflammatory disease or graft versus host disease. RRV's are capable of surviving inactivation in human serum thereby allowing efficient gene transfer over prolonged periods of time.

(Updated on 25-MAR-2003 to correct PI field.)

Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;

Best Local Similarity 99.9%; Freq. No. 0;

Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATRRYLGAVELSWDYNQSDGLGELPVDARFPVRPKSPFNNTSVYKTLFVEFTVHLEN 60

20 ATRRYLGAVELSWDYNQSDGLGELPVDARFPVRPKSPFNNTSVYKTLFVEFTVHLEN 79

61 IAKPRPPMGLLGTQAEVYDVITLKNASHPVLSLHAGVSVWKASEGAEYDDQTSQ 120

Db

80 TAKPRPPMGLLGTQAEVYDVITLKNASHPVLSLHAGVSVWKASEGAEYDDQTSQ 139

QY

121 REKEDDKVFPGGSHYVQVVKENGPMASDPLCLTYSLSHVDVLKOLNSGLIGALLVCR 180

Db

140 REKEDDKVFPGGSHYVQVVKENGPMASDPLCLTYSLSHVDVLKOLNSGLIGALLVCR 199

QY

181 EGS�AKBKQTQLHKFILLFAVFDGKSWHSEPKNSLMQDRDAASARAWPKMHTVNGVNR 240

Db

200 EGS�AKBKQTQLHKFILLFAVFDGKSWHSEPKNSLMQDRDAASARAWPKMHTVNGVNR 259

QY

241 SLPGLIGCHRSYVWHVIGMGTTPVHSTFLECHTFLVNRHQASLEISPTFLTAQTLL 300

Db

260 SLPGLIGCHRSYVWHVIGMGTTPVHSTFLECHTFLVNRHQASLEISPTFLTAQTLL 319

QY

301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPQPMKKNNEAEYDDDLTDSMDVVRF 360

Db

320 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPQPMKKNNEAEYDDDLTDSMDVVRF 379

QY

361 DDNSPSFIQIRSAVAKHPKTVVHYIAAEBEDWDYAPLVLPAPDRRSYKSYQLNNGPQIG 420

Db

380 DDNSPSFIQIRSAVAKHPKTVVHYIAAEBEDWDYAPLVLPAPDRRSYKSYQLNNGPQIG 439

QY

421 RYKVKVFMAYTDETFKTREAIQHESGILGPLLYGEVGDFTLLIEKQASRPYNIYPHGI 480

Db

440 RYKVKVFMAYTDETFKTREAIQHESGILGPLLYGEVGDFTLLIEKQASRPYNIYPHGI 499

QY

481 TDVREPLYSRRLPKGVKHLKDFPILFGEIFPKYKWTVTVDGPKSPRCLTRYSSFVNME 540

Db

500 TDVREPLYSRRLPKGVKHLKDFPILFGEIFPKYKWTVTVDGPKSPRCLTRYSSFVNME 559

QY

541 RDLASGLTGPLLI CYKESVDQNGQINSDKNVILFSVFDNRSWYLTENQRFPLNPAG 600

Db

560 RDLASGLTGPLLI CYKESVDQNGQINSDKNVILFSVFDNRSWYLTENQRFPLNPAG 619

QY

601 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSGATDFLSVFFSGVTGKH 660

Db

620 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSGATDFLSVFFSGVTGKH 679

QY

661 KMVYEDTTLTPFPFGSETVFMENPGLMILGCHNSDFRNRGMTALLKVSCKDKTGDYE 720

Db

680 KMVYEDTTLTPFPFGSETVFMENPGLMILGCHNSDFRNRGMTALLKVSCKDKTGDYE 739

QY

721 DSYEDISAYLSKNAIEPRFSQNPVLPKHORITRTTLOSDEEIDYDDTISVEMKK 780

Db

740 DSYEDISAYLSKNAIEPRFSQNPVLPKHORITRTTLOSDEEIDYDDTISVEMKK 799

QY

781 EDFDIYDEDNOSPRSFQKTRHYFIAAVERLDYGMSSPHVLNRNRAQSGSVPOFKVY 840

Db

800 EDFDIYDEDNOSPRSFQKTRHYFIAAVERLDYGMSSPHVLNRNRAQSGSVPOFKVY 859

QY

841 FQEFDTGSGFTQPLYRGELNEHLGLGPYIRAEVDNIMVTFRNOASRPYSFSSLIYEE 900

Db

860 FQEFDTGSGFTQPLYRGELNEHLGLGPYIRAEVDNIMVTFRNOASRPYSFSSLIYEE 919

QY

901 DQOQAEPRKPKVFNKTKTVFKVQHMATPKDEFCDCKAWAYFSDVDLEKXVHSLGTP 960

Db

920 DQOQAEPRKPKVFNKTKTVFKVQHMATPKDEFCDCKAWAYFSDVDLEKXVHSLGTP 979

QY

961 LLVCHTNTLNPAHQKQVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNTQMEDPTKE 1020

Db

980 LLVCHTNTLNPAHQKQVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNTQMEDPTKE 1039

QY

1021 NYRHAINGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIHPSGHVFTVRKKEEYQOA 1080

Db

1040 NYRHAINGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIHPSGHVFTVRKKEEYQOA 1099

QY

1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHGHMSTFLVYSNKCTQPLGMASGHIRD 1140

Db

1100 LYNLYPGVFETVEMLPKAGIWRVECLIGELHGHMSTFLVYSNKCTQPLGMASGHIRD 1159

QY

1141 FQITASQGYQWAPKLAFLPHYSINAMSTKEPFSWIKVDLLAPMIHGIKTOGARQKFS 1200

Db 1160 FQTASQYQCMAPKARLHLYSGSINAWSTKPFSSWIKWDLAPMIHGIKTQCARQKFS 1219
 QY 1201 SLYISQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGICKINFPPIIARVIRLHPT 1260
 Db 1220 SLYISQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGICKINFPPIIARVIRLHPT 1279
 QY 1261 HYSIRSTLRMLGCDLNSCMBPLGMSKASISDAQITASSYFTNMFATWSPSKARLHLOQ 1320
 Db 1280 HYSIRSTLRMLGCDLNSCMBPLGMSKASISDAQITASSYFTNMFATWSPSKARLHLOQ 1339
 QY 1321 RSNARPOVNNPKENLQVDFQKTKMKTGVTGQVKSLLTSMYKKEFLISSQDGHOWTLF 1380
 Db 1340 RSNARPOVNNPKENLQVDFQKTKMKTGVTGQVKSLLTSMYKKEFLISSQDGHOWTLF 1399
 QY 1381 FQNGKVKVFGNQDSFTPVVNSLDPPLITRLYRIHQSWHQIALRMEVLGCEAQQDLY 1438
 Db 1400 FQNGKVKVFGNQDSFTPVVNSLDPPLITRLYRIHQSWHQIALRMEVLGCEAQQDLY 1457

RESULT 4

ID AAW44372

XX AAW44372 standard; Protein; 1457 AA.

AC AAW44372;

XX 20-JUL-1998 (first entry)

XX Human Factor VIII SQN deletion mutant.

DE Factor VIII; blood clotting; haemophilia A; gene therapy;
 KW retrovirus; vector; human.

XX Homo sapiens.

OS Synthetic.

PN WO9800542-A2.

XX 08-JAN-1998.

XX 02-JUL-1997; 97MO-US11785.

XX 04-JUN-1997; 97US-0869309.

PR 03-JUL-1996; 96US-0645601.

PR 13-AUG-1996; 96US-0696381.

XX (CHIR) CHIRON CORP.

XX Allen JR, Barber JR, Boder M, Chang SM, Chong K;

PI De La Vega D, Depolo NJ, Greengard J, Hsu DC, Ibanez CE;

PI Jolly DJ, Mittelstaedt DM, Prussak CE, Respass JG;

XX WPI; 1998-086967/08.

DR N-PSDB; AAV15338.

XX New replication defective recombinant retroviruses - which express B

PT domain-deleted human factor VIII or human factor IX for the

PT treatment of haemophilia

XX Claim 5; Page 175-180; 236pp; English.

XX This polypeptide comprises the B domain deletion mutant SQN of

CC human Factor VIII. The SQN mutant is created by fusing Ser-743 to

CC Gln-1638 of native Factor VIII (see AAW44372) to form a Ser-Gln-Aan

CC (SQN) link between the A2 and A3 factor VIII domains. A DNA

CC sequence encoding the SQN deletion mutant is provided in AAV15338.

CC When compared to plasmid-derived Factor VIII, the SQN deletion does

CC not influence the in vivo pharmacokinetics, but the reduced size of

CC the molecule appears to decrease proteolytic degradation. The

CC invention relates to preparations of replication defective

CC recombinant retrovirus (RV) expressing a B domain-deleted human

CC Factor VIII protein, where the recombinant RV is capable of

CC infecting human cells, is resistant to degradation by human

CC complement and is capable of inducing long-term (at least 30 days

CC and up to 6 months or longer post-injection) systemic expression of
 CC Factor VIII when administered to a haemophilia A patient.

SQ Sequence 1457 AA;

Query Match 99.9%; Score 7684; DB 19; Length 1457;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATRRYVLGAVELSDWYMQSDIGELPVDARFPFRVPKSPFPNTSVYKKTFLVFETVHLFN 60
 Db 20 ATRRYVLGAVELSDWYMQSDIGELPVDARFPFRVPKSPFPNTSVYKKTFLVFETVHLFN 79
 QY 61 IAKPRPPMGLGPTTQAEVYDTVTITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 120
 Db 80 IAKPRPPMGLGPTTQAEVYDTVTITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQ 139
 QY 121 RKEKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 Db 140 RKEKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
 QY 181 EGS�AKEKTOTLHKFTILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTVNGYVNR 240
 Db 200 EGS�AKEKTOTLHKFTILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTVNGYVNR 259
 QY 241 SLFGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTPLVNRHROASLEISPIITFLTAQTLL 300
 Db 260 SLFGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTPLVNRHROASLEISPIITFLTAQTLL 319
 QY 301 MDLQGPLLPCHISSHQHDGMEAYVVDSCPEPQJRMKNNEAEVDLDLTSEMDVVRP 360
 Db 320 MDLQGPLLPCHISSHQHDGMEAYVVDSCPEPQJRMKNNEAEVDLDLTSEMDVVRP 379
 QY 361 DDNSPSPFIQIRSVAKGHPKTVWHYIAAEEEDWDYAPLVLPDDRYSKSOYLNGEQRIG 420
 Db 380 DDNSPSPFIQIRSVAKGHPKTVWHYIAAEEEDWDYAPLVLPDDRYSKSOYLNGEQRIG 439
 QY 421 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEGVGDTLIIIFKQASRPVNTYPHGI 480
 Db 440 RYKVKVRFMAYTDETFKTREAIQHESGILGPLLYGEGVGDTLIIIFKQASRPVNTYPHGI 499
 QY 481 TVRPLYSRRLPKGVKHLKDPFILPGEIFKFKWTVTVEDGPTKSPRCLTRYYSFVNME 540
 Db 500 TVRPLYSRRLPKGVKHLKDPFILPGEIFKFKWTVTVEDGPTKSPRCLTRYYSFVNME 559
 QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVDENRSHVLTENIORLPNAG 600
 Db 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVDENRSHVLTENIORLPNAG 619
 QY 601 VQLEDPEFQASNMHSINGYVFDLSIQLSVCLHEVAYWILSICAQTDPLSVFSGYTFKH 660
 Db 620 VQLEDPEFQASNMHSINGYVFDLSIQLSVCLHEVAYWILSICAQTDPLSVFSGYTFKH 679
 QY 661 KMYVEDTITLTPFSGETVFMENPGLMILGCHNSDFRNGMTALLKVVSSDKNTGDYYE 720
 Db 680 KMYVEDTITLTPFSGETVFMENPGLMILGCHNSDFRNGMTALLKVVSSDKNTGDYYE 739
 QY 721 DSYEDISAYLLSKNNAIEPRSPVLPKHHQREITRTTLQSDQEEIDYDDTISVENMKK 780
 Db 740 DSYEDISAYLLSKNNAIEPRSPVLPKHHQREITRTTLQSDQEEIDYDDTISVENMKK 799
 QY 781 EDFDIYDEBNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQFKKV 840
 Db 800 EDFDIYDEBNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQFKKV 859
 QY 841 FOETDGSFTQPLVRGELNEHLGLLGPYIRAEVDNINMTVFRNQASRPYSYSLISYEE 900
 Db 860 FOETDGSFTQPLVRGELNEHLGLLGPYIRAEVDNINMTVFRNQASRPYSYSLISYEE 919
 QY 901 DQOGAPRKNFVKPNETKTYFWKQHHMPTKDEFCCKAWAYFSDVDLEKDVHSGLIGP 960
 Db 920 DQOGAPRKNFVKPNETKTYFWKQHHMPTKDEFCCKAWAYFSDVDLEKDVHSGLIGP 979

Qy 961 LLVCHNTNLPAHGRQVTVQEPALFFTFIDETKSWYFTENWENCRAPCNQIOMEDPTFXE 1020
 Dd 980 LLVCHNTNLPAHGRQVTVQEPALFFTFIDETKSWYFTENWENCRAPCNQIOMEDPTFXE 1039
 Qy 1021 NYRFAHNGYIMDTPLGLMAQDORIRWYLLSGNSNENIHSIHFSGHVFTVRKERYKMA 1080
 Dd 1040 NYRFAHNGYIMDTPLGLMAQDORIRWYLLSGNSNENIHSIHFSGHVFTVRKERYKMA 1099
 Qy 1081 LYNLYPGVFTVEMLPKAGINRVECLIGBHLHAGMSTLFLVYSNKCQTPGLGASGHIRD 1140
 Dd 1100 LYNLYPGVFTVEMLPKAGINRVECLIGBHLHAGMSTLFLVYSNKCQTPGLGASGHIRD 1159
 Qy 1141 FQITASGOVGMAPKLARLHYSGINAWSTKEPFSMIKVDLLAPMIHGHKTQARQKPS 1200
 Dd 1160 FQITASGOVGMAPKLARLHYSGINAWSTKEPFSMIKVDLLAPMIHGHKTQARQKPS 1219
 Qy 1201 SLYISQFIIMYSIDGKKWQYRGNSGTGLMVFFGNVDSSGKININPPIIARYIRLHPT 1260
 Dd 1220 SLYISQFIIMYSIDGKKWQYRGNSGTGLMVFFGNVDSSGKININPPIIARYIRLHPT 1279
 Qy 1261 HYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARHLQ 1320
 Dd 1280 HYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARHLQ 1339
 Qy 1321 RSNARPOVNNPKEWLQVDPQKTMKVTGTTQGVKSILTSMYKBEFLISSQDGHQWTLF 1380
 Dd 1340 RSNARPOVNNPKEWLQVDPQKTMKVTGTTQGVKSILTSMYKBEFLISSQDGHQWTLF 1399
 Qy 1381 FQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRIHFQSWVHQIALRMEVLGCEAODLY 1438
 Dd 1400 FQNGKVKVFGQNDSTFPVNSLDPPLLTRYLRIHFQSWVHQIALRMEVLGCEAODLY 1457

RESULT 5

AAW23414
 ID AAW23414 standard; Protein; 1471 AA.

AC AAW23414;

XX 08-APR-1998 (first entry)

DE Human B-domain deleted factor VIII protein.

XX Post-translational regulatory element; PRE; enhancer II; intronless gene;
 KW surface antigen gene; cytoplasmic accumulation; targeted delivery;
 KW near consensus splice sequence; blood coagulation factor; factor VIII;
 KW factor IX.

XX Homo sapiens.

XX W09733994-A1.

XX 18-SEP-1997.

PF 10-MAR-1997; 97WO-0503561.

XX 11-MAR-1996; 96US-0683839.

PR (IMMU-) IMMUNE RESPONSE CORP.

PA Bidlingmaier S, Ill CR;

XX WPI; 1997-470874/43.

DR N-PSDB; AAT73164.

XX Vector for increased expression of intronless genes - comprises
 PT intronless gene with at least one near consensus splice sequence, a
 PT promoter and at least one viral cis-acting post-transcriptional
 PT regulatory element

PS Example 1; Pages 31-36; 59pp; English.

XX The present sequence represents human B-domain deleted factor VIII

CC The cDNA encoding this protein also contains, 3' of the coding region, a
 CC post-translational regulatory element (PRE) of the Hepatitis B virus.
 CC PRE sequences have been shown to function in cis to increase the
 CC steady-state levels of surface gene transcripts by facilitating
 CC cytoplasmic accumulation of these transcripts. The above nucleic acid
 CC sequence is part of a novel vector, comprising an intronless gene
 CC containing 1 or more near consensus splice sequences operably linked to
 CC a promoter sequence so that the gene is transcribed in a cell.
 CC Intronless gene transcripts which contain near consensus splice site
 CC sequences are believed to get tied up in the nucleus of the cell where
 CC splicing occurs, rather than being transported to the cytoplasm where
 CC they can be translated into proteins. The PRE sequences are transcribed
 CC along with the gene, causing export of the gene transcript from the
 CC nucleus into the cytoplasm of the cell. The vector can be used
 CC to increase the expression of an intronless gene containing at least one
 CC near consensus splice sites, preferably cDNA encoding a blood coagulation
 CC factor, particularly factor VIII or IX. The complex allows the targeted
 CC delivery of the vector to a specific cell, e.g. hepatocytes when the
 CC ligand is an asialoglycoprotein which binds the asialoglycoprotein
 CC receptor present on their surface.

XX Sequence 1471 AA;

Query Match 99.8%; Score 7674; DB 18; Length 1471;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

Qy 1 ATRRYVLGAVELSDMYQMSDLGELPVDARFPFRVPKSPFPNTSVVYKKTFLVEFTVHLFN 60
 Dd 20 ATRRYVLGAVELSDMYQMSDLGELPVDARFPFRVPKSPFPNTSVVYKKTFLVEFTVHLFN 79
 Qy 61 IAKPRPPWMLLGPITQAEVDTVTITLKNMASHPVSLHAGVSVTWKASEGAEYDDQTSQ 120
 Dd 80 IAKPRPPWMLLGPITQAEVDTVTITLKNMASHPVSLHAGVSVTWKASEGAEYDDQTSQ 139
 Qy 121 REKEDDKVFPFGSSHTYVWQVLKENGPMASDPLCLTVSYLVSHVDVAKDLSGLIGALLVCR 180
 Dd 140 REKEDDKVFPFGSSHTYVWQVLKENGPMASDPLCLTVSYLVSHVDVAKDLSGLIGALLVCR 199
 Qy 181 EGS�AKEKTQTLHKFILLFAVDFEGKSWHSEKNSLMQDRDAASARAPKQHTVGVYNR 240
 Dd 200 EGS�AKEKTQTLHKFILLFAVDFEGKSWHSEKNSLMQDRDAASARAPKQHTVGVYNR 259
 Qy 241 SLFGLIGCHRKSVYWHVTCMGTTTPVHSIFLGHTFLVRNHRQASLEISPTFLFAQTLL 300
 Dd 260 SLFGLIGCHRKSVYWHVTCMGTTTPVHSIFLGHTFLVRNHRQASLEISPTFLFAQTLL 319
 Qy 301 MDLGQFLFCHISSHQHDMGEAYVKVDCSPBEPQLRMKNNEABDYDDDLTDSMDVVRP 360
 Dd 320 MDLGQFLFCHISSHQHDMGEAYVKVDCSPBEPQLRMKNNEABDYDDDLTDSMDVVRP 379
 Qy 361 DDNSPSTQIRSVAKGHPKTVWHVYIAAEEEDMDVAPLVAPDDDSYKSYQYLNNGFORIG 420
 Dd 380 DDNSPSTQIRSVAKGHPKTVWHVYIAAEEEDMDVAPLVAPDDDSYKSYQYLNNGFORIG 439
 Qy 421 RXYKKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNYIPGCI 480
 Dd 440 RXYKKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNYIPGCI 499
 Qy 481 TDVRELYSRRLPKGVKHLKDFPILFGEIFKYNKWTYVVEDGPTKSDPRCLTRYSSFVNME 540
 Dd 500 TDVRELYSRRLPKGVKHLKDFPILFGEIFKYNKWTYVVEDGPTKSDPRCLTRYSSFVNME 559
 Qy 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNWILFSVFDENRSHWLTENTQRFPLNPAG 600
 Dd 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNWILFSVFDENRSHWLTENTQRFPLNPAG 619
 Qy 601 VQLEDPFQASNMHNSINGVYVDFSLSQVCLHEVAYWYLSIGACTDPLSFSSGYTPKH 660
 Dd 620 VQLEDPFQASNMHNSINGVYVDFSLSQVCLHEVAYWYLSIGACTDPLSFSSGYTPKH 679
 Qy 661 KXVYEDTTLTLPFFSGETVFMSENFGMLWILGHNSDFNRGWTALLKVSSCDKNTGYYE 720

Db 680 KMVYEDTLTLPFSGETVFMSENPGILWILGCHNSDFRNMGNTALLKVSQCDKNTGDYVE 739
 QY 721 DSYEDISAYLLSKNNAIEPRSFQN-----PPVLKRHQREITRTTLOSQOE 766
 Db 740 DSYEDISAYLLSKNNAIEPRSFQNSHSTROKQFNATPPVLRHQREITRTTLOSQOE 799
 QY 767 EIDYDDTISVEMKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLRN 826
 Db 800 EIDYDDTISVEMKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLRN 859
 QY 827 RAQSGSVPOKPKVFOEFTDGSSTOPLYRGELNEHLGLLPYIRAEVDNIMWTFNQAS 886
 Db 860 RAQSGSVPOKPKVFOEFTDGSSTOPLYRGELNEHLGLLPYIRAEVDNIMWTFNQAS 919
 QY 887 RPYSFYSSLIYSBEDQORQAEPRKFNFKNETKTYFKVQOHMAPTKDFDCKAWAYFSD 946
 Db 920 RPYSFYSSLIYSBEDQORQAEPRKFNFKNETKTYFKVQOHMAPTKDFDCKAWAYFSD 979
 QY 947 VLEKDVHSLIGLPLLVCHTNTLNPAHGQVTVQEFALPFTTFDETKSWYFTENMERNCR 1006
 Db 980 VLEKDVHSLIGLPLLVCHTNTLNPAHGQVTVQEFALPFTTFDETKSWYFTENMERNCR 1039
 QY 1007 APCNIOWEPTFKENYRFAINCYINDTLPGLYMAQDQIRWYLLSGNSNENIHSIHFG 1066
 Db 1040 APCNIOWEPTFKENYRFAINCYINDTLPGLYMAQDQIRWYLLSGNSNENIHSIHFG 1099
 QY 1067 HVFTVRKKEEYKMALYNLYPGVETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1126
 Db 1100 HVFTVRKKEEYKMALYNLYPGVETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNK 1159
 QY 1127 CQTPLGMAHGHRDFOITASGQYQWAPKPLARLHYSGSINAWSTKPPFSWIKVDLLAPMI 1186
 Db 1160 CQTPLGMAHGHRDFOITASGQYQWAPKPLARLHYSGSINAWSTKPPFSWIKVDLLAPMI 1219
 QY 1187 IHGIKTQGARQKPFSSLYISQFIIMYSLDGKMTYRGNSTGTLMVFFGNVDSSGKHNF 1246
 Db 1220 IHGIKTQGARQKPFSSLYISQFIIMYSLDGKMTYRGNSTGTLMVFFGNVDSSGKHNF 1279
 QY 1247 NPPIIARYIRLHPHTHYSIIRSTLWELMGCDLNSCMLPGWESKAISDAITASSYFTNMF 1306
 Db 1280 NPPIIARYIRLHPHTHYSIIRSTLWELMGCDLNSCMLPGWESKAISDAITASSYFTNMF 1339
 QY 1307 ATWSPSKARLHLQGRSNARWPQVNNPKELQVDFOKTMKVTGTTQGVKSLTSMYVKEF 1366
 Db 1340 ATWSPSKARLHLQGRSNARWPQVNNPKELQVDFOKTMKVTGTTQGVKSLTSMYVKEF 1399
 QY 1367 LISSQDGHQWTLFFQNGKVKYFQGNQDFTPVNSLDPLLTLYLRHIFQSWHQAIALR 1426
 Db 1400 LISSQDGHQWTLFFQNGKVKYFQGNQDFTPVNSLDPLLTLYLRHIFQSWHQAIALR 1459
 QY 1427 MEVLGCEAQDLY 1438
 Db 1460 MEVLGCEAQDLY 1471

RESULT 6

AAB67959
 ID AAB67959 standard; Protein; 1471 AA.

AC AAB67959;

DT 29-JUN-2001 (first entry)

DE Amino acid sequence of human B-domain deleted factor VIII.

KW Adeno-associated virus vector; B-domain; factor VIII; haemophilia A;

OS coagulation disorder.

QX Homo sapiens.

XX WO200127303-A1.

PD 19-APR-2001.

XX 12-OCT-2000; 2000WO-US28221.
 XX 12-OCT-1999; 99US-0158780.
 XX (UYNC-) UNIV NORTH CAROLINA.
 PA Walsh CE, Chao H, Burstein H, Lynch CM, Stepan AM, Munson K;
 PI WPI; 2001-273781/28.
 DR N-PSDB; AAF84647.
 XX
 PT New recombinant adeno-associated virus vector, useful for treating
 PT haemophilia A, comprises heterologous nucleotide sequence encoding
 PT B-domain deleted human factor VIII operably linked with liver-preferred
 PT expression control element -
 XX
 PS Example 1; Fig 1; 87pp; English.
 CC The specification describes a recombinant adeno-associated virus (rAAV)
 CC vector. The vector comprises a heterologous nucleotide sequence
 CC encoding B-domain deleted factor VIII operably linked with at least one
 CC enhancer and at least one promoter. The method results in the production
 CC of high titer rAAV vector stocks carrying the B-domain deleted factor
 CC VIII transgene and expression cassettes, which generate adequate titers
 CC of virus for in vivo administration. The recombinant vectors are useful
 CC for treating haemophilia A, where the liver expresses the encoded
 CC B-domain deleted factor VIII, which is secreted into the blood. They are
 CC also useful for the treatment of other coagulation disorders. The
 CC present sequence represents a B-domain deleted factor VIII.
 XX Sequence 1471 AA;

Query Match 99.8%; Score 7674; DB 22; Length 1471;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

QY 1 ATRRYLGAVELSWDMQSDLGELPVDARFPFRPKSPFNSTVVKKTLFVEFTVHLFN 60
 Db 20 ATRRYLGAVELSWDMQSDLGELPVDARFPFRPKSPFNSTVVKKTLFVEFTVHLFN 79
 QY 61 TAKPPPPMGLGPTTQAEVYDVTVITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 120
 Db 80 TAKPPPPMGLGPTTQAEVYDVTVITLKNASHPVSLHAGVSVKASEGAEYDDQTSQ 139
 QY 121 REKEDDKVFPGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLKDLNSGLIGALLVCR 180
 Db 140 REKEDDKVFPGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLKDLNSGLIGALLVCR 199
 QY 181 EGSIAKEKTQTLHKPILLFAVFEDEKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240
 Db 200 EGSIAKEKTQTLHKPILLFAVFEDEKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 259
 QY 241 SLPLGLIGCHRSVYVHWIGMGTTPVHSIFLEHTFLVRNHRQASLEISPTITFLTAQTL 300
 Db 260 SLPLGLIGCHRSVYVHWIGMGTTPVHSIFLEHTFLVRNHRQASLEISPTITFLTAQTL 319
 QY 301 MDLGOFLFCHISSHQHDGMEAYVKVDSCEPQRLMKNEEAEYDDDLTDSMDVVRP 360
 Db 320 MDLGOFLFCHISSHQHDGMEAYVKVDSCEPQRLMKNEEAEYDDDLTDSMDVVRP 379
 QY 361 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSKYSQVNLNNGPKRIG 420
 Db 380 DDNSPSFIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDRSKYSQVNLNNGPKRIG 439
 QY 421 RYKVKVRFMAYTDTFTKTREAIQHESGILGPLLYGEGVDTLLIIFKQASRPYNIYPHGI 480
 Db 440 RYKVKVRFMAYTDTFTKTREAIQHESGILGPLLYGEGVDTLLIIFKQASRPYNIYPHGI 499
 QY 481 TDVRLPLYSRRLPKGVKHLKOPFILLPGEIKFYKWTVTVEDGPKTCKPCLTRYYSFVNME 540
 Db 500 TDVRLPLYSRRLPKGVKHLKOPFILLPGEIKFYKWTVTVEDGPKTCKPCLTRYYSFVNME 559

QY 541 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNWILFSVDFENRSWYLTENIQRFLENPAG 600
DB 560 RDLASGLIGPLLCYKESVDQRGNQIMSDKRNWILFSVDFENRSWYLTENIQRFLENPAG 619
QY 601 VOLEDEPFOASNIWHSINGVVFDSQLSVCLHEVAVWYILSIGAQTDELSPVSGYTFKH 660
DB 620 VOLEDEPFOASNIWHSINGVVFDSQLSVCLHEVAVWYILSIGAQTDELSPVSGYTFKH 679
QY 661 KMVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSQCDKNTGDYIE 720
DB 680 KMVEDTLTLFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSQCDKNTGDYIE 739
QY 721 DSYEDISAYLLSKNAIEPRFSQN-----PPVLKRHOREITRITLOSQOE 766
DB 740 DSYEDISAYLLSKNAIEPRFSQNSHPSTROKQFNATPPVLKRHOREITRITLOSQOE 799
QY 767 EIDYDDTISVEMKEDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLRN 826
DB 800 EIDYDDTISVEMKEDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLRN 859
QY 827 RAQSGSVQPKVQFQFTDGSFTQPLYRGELNEHLGLLOPYTRAEDVEDNIMVTFRNQAS 886
DB 860 RAQSGSVQPKVQFQFTDGSFTQPLYRGELNEHLGLLOPYTRAEDVEDNIMVTFRNQAS 919
QY 887 RPYSFYSSLSYEDQOQAPRKNFVKPNETKTYFWKVQHHMAPTKDEPCKAWAYFSD 946
DB 920 RPYSFYSSLSYEDQOQAPRKNFVKPNETKTYFWKVQHHMAPTKDEPCKAWAYFSD 979
QY 947 VDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCR 1006
DB 980 VDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTTIDETKSWYFTENMERNCR 1039
QY 1007 APCNIQMEDTFKENTYRFAHNGYIMDTPLGLWAQDQRTWYLLSGNSNETHSHFSG 1066
DB 1040 APCNIQMEDTFKENTYRFAHNGYIMDTPLGLWAQDQRTWYLLSGNSNETHSHFSG 1099
QY 1067 HVFTVRKKEVKMALYNLYPGVETVEMLPKAGIWECLIGELHAGHAGSTLFLVYSNK 1126
DB 1100 HVFTVRKKEVKMALYNLYPGVETVEMLPKAGIWECLIGELHAGHAGSTLFLVYSNK 1159
QY 1127 CQPLGASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKPFWSIKVDLLAPMI 1186
DB 1160 CQPLGASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKPFWSIKVDLLAPMI 1219
QY 1187 IHGKTQAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTCTLMVFGNVDSSGIXHNIF 1246
DB 1220 IHGKTQAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTCTLMVFGNVDSSGIXHNIF 1279
QY 1247 NPPIIARYIRLHPHYTHYSIRSTRMELMGCGLNSCSMPLGNSKASDAQITASSYFTNMF 1306
DB 1280 NPPIIARYIRLHPHYTHYSIRSTRMELMGCGLNSCSMPLGNSKASDAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHLOGRSNARPOVNNPKWLOVDFOKTMKVTGTTQGVKSLLSMTYKEP 1366
DB 1340 ATWSPSKARLHLOGRSNARPOVNNPKWLOVDFOKTMKVTGTTQGVKSLLSMTYKEP 1399
QY 1367 LISSQDGHQWTLFPFGKGVKVFQGNQDSFTPVVNSLDPPLLTRYLRHPQSWHOTALR 1426
DB 1400 LISSQDGHQWTLFPFGKGVKVFQGNQDSFTPVVNSLDPPLLTRYLRHPQSWHOTALR 1459
QY 1427 MEVLGCEAQDLY 1438
DB 1460 MEVLGCEAQDLY 1471

RESULT 7

AAR12971

ID AAR12971 standard; protein; 1440 AA.

XX AC

XX AC

XX AC

DT 25-MAR-2003 (updated)

DT 09--JAN-2003 (updated)

DT 02-OCT-1991 (first entry)
XX Factor VIII:SQ.
DE Factor VIII:SQ.
XX Factor VIII; B domain; haemophilia.
KW Factor VIII; B domain; haemophilia.
XX Unidentified.
OS Unidentified.
XX WO9109122-A.
PN 27-JUN-1991.
XX 06-DEC-1990; 90WO-SE00809.
PP 15-DEC-1989; 89SE-0004239.
XX (KABI) KABIVITRUM AB.
PA Almstedt AB, Hellstrom EM, Larsson K, Lind P, Sandberg HI;
PI Spira J, Sydowback MW;
PI WPI; 1991-208148/28.
XX Recombinant human factor VIII deriv. deoxyribonucleic acid -
PT encoding protein comprising two chains linked by segment of B domain.
XX Disclosure; Fig 1; 35pp; English.
XX The protein is a fusion between Phe 742 and Ser 1637 of the factor
CC VIII protein (factor VIII:SQ). In order to produce a
CC factor VIII deletion derivative that can be produced in vivo and/or
CC in vitro, to a two chain protein consisting of polypeptide chains of
CC 90 kD and 80 kD, the amino acid sequences surrounding Arg 740 and
CC Arg 1648 have to be conserved in order to preserve the structural
CC requirements for correct cleavage. In this example, amino acids 743
CC to 1636 of the full-length factor VIII polypeptide are deleted. A
CC new polypeptide chain is obt. where there are 14 amino acids
CC linking Arg 740 and Arg 1648. Of these 14 amino acids, the sequence
CC of the five N-terminal ones directly corresponds to the five amino
CC acids following Arg 740 in full-length factor VIII. Also, the sequence
CC of the 12 C-terminal amino acids of the above 14 amino acids fragment
CC directly corresponds to the 12 amino acids preceding Glu 1689 in full-
CC length factor VIII, thus creating a 3 amino acid overlap between the
CC N- and C-terminal regions of the B-domain.
CC The factor VIII deriv. is useful for treating haemophilia or
CC haemophilia A. It has the biological characteristics of plasma derived
CC factor VIII.
CC In order to index this example, the factor VIII:QD amino acid
CC sequence was retrieved from WO8800831 (AAP80265).
CC The amino acid numbering in the above comments is reproduced from the
CC fig. description in the specification. Note that Arg 740 is Arg 742
CC in the indexed sequence, etc. Also, Asn 745 (N-terminal link overlap)
CC is Asp 747 in AAP80265, but indexed as Asn to reproduce the fusion
CC fragment as shown in fig.1, and Asn 745 (C-terminal link overlap).
CC (Updated on 09-JAN-2003 to add missing OS field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX Sequence 1440 AA;
SQ Query Match 99.6%; Score 7659.5; DB 12; Length 1440;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1434; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 1 ATRRYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFPNTSVYKTLFVEFTVHLEN 60
DB 2 ATRRYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFPNTSVYKTLFVEFTVHLEN 61
QY 61 IAKPRPPMGLGPTIQAEVYDVTITLKNMASHPVSLHAGVSVTWKSEGREYDDQTSQ 120
DB 62 IAKPRPPMGLGPTIQAEVYDVTITLKNMASHPVSLHAGVSVTWKSEGREYDDQTSQ 121
QY 121 REKEDDKVPFGSGSYHYVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGIGALLVCR 180

Db 122 REKEDDKVFGSGSHYVWVQVLKENGPMASDPLCLTYSLVSHVDLVKDLNSGLIGALLVCR 181
QY 181 EGS LAKEKQTTLHKFTLLFAVPDEGKSHSETKNSLMQDRDAASARAWPKMTVYVNR 240
Db 182 EGS LAKEKHTTLHKFTLLFAVPDEGKSHSETKNSLMQDRDAASARAWPKMTVYVNR 241
QY 241 S-LPGLIGCHRSVYVHVIGMGTTPEVHSIFLEGHTFLVNRHQASLEISPTITFAQTL 299
Db 242 SLLPGLIGCHRSVYVHVIGMGTTPEVHSIFLEGHTFLVNRHQASLEISPTITFAQTL 301
QY 300 LMDLGQFLFCHISSHQHGMAYVYKVDSCPEPQLRMKNBEAEDYDDDLTDSMDVVR 359
Db 302 LMDLGQFLFCHISSHQHGMAYVYKVDSCPEPQLRMKNBEAEDYDDDLTDSMDVVR 361
QY 360 FDDNDSPSFIQRSVAKPKTWVHYIAAEEDMDYAPLAPDDRSYKSYLVNNGPQRI 419
Db 362 FDDNDSPSFIQRSVAKPKTWVHYIAAEEDMDYAPLAPDDRSYKSYLVNNGPQRI 421
QY 420 GRKYKVRMAYTDEFTKTRAIQHSIGILGPLYGEVGDTLIIIPKQASRPYNIYPHG 479
Db 422 GRKYKVRMAYTDEFTKTRAIQHSIGILGPLYGEVGDTLIIIPKQASRPYNIYPHG 481
QY 480 ITDVRPLYSRRLPKGVKHLKDPILPGBIEFKYKWTTVEDGPKSDPRCLTRYISFVNM 539
Db 482 ITDVRPLYSRRLPKGVKHLKDPILPGBIEFKYKWTTVEDGPKSDPRCLTRYISFVNM 541
QY 540 ERDLASGLIGPLLICVYESVDQGNQIMSDKRNILFVSFVDENRSHYLTENIQRFLNPA 599
Db 542 ERDLASGLIGPLLICVYESVDQGNQIMSDKRNILFVSFVDENRSHYLTENIQRFLNPA 601
QY 600 GVOLEDPFOASNIMHSINGYVDFSLQSLVCLHEVAYWYILSIGAQTDFLSVFFSGYTFK 659
Db 602 GVOLEDPFOASNIMHSINGYVDFSLQSLVCLHEVAYWYILSIGAQTDFLSVFFSGYTFK 661
QY 660 HKWVYEDTTLTPFSGSETVMSNENPCLMILGCHNSDFNRGHTALLKVCSCDKMTGDIY 719
Db 662 HKWVYEDTTLTPFSGSETVMSNENPCLMILGCHNSDFNRGHTALLKVCSCDKMTGDIY 721
QY 720 EDSYEDISAYLLSKNNAIEPRFSQNPVLPVKKHOREITRTTLQSDOEEDYDDTISVENK 779
Db 722 EDSYEDISAYLLSKNNAIEPRFSQNPVLPVKKHOREITRTTLQSDOEEDYDDTISVENK 781
QY 780 KEDFDIYDEDNOSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNRAQSGSVFQFKV 839
Db 782 KEDFDIYDEDNOSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNRAQSGSVFQFKV 841
QY 840 VQOEFTDGSFTOPLVYRGELNEHLGLGPYIRAEVEDNIMVTFNRQASRPYSYSLISYE 899
Db 842 VQOEFTDGSFTOPLVYRGELNEHLGLGPYIRAEVEDNIMVTFNRQASRPYSYSLISYE 901
QY 900 EDQRCQAEPRKXNFVKNETKTYFWKYQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIG 959
Db 902 EDQRCQAEPRKXNFVKNETKTYFWKYQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIG 961
QY 960 PLLVCHTNTLNAHGROVTVQFALPFTTDFETKSWYFTENMERNCRAPCNQIMBPTFK 1019
Db 962 PLLVCHTNTLNAHGROVTVQFALPFTTDFETKSWYFTENMERNCRAPCNQIMBPTFK 1021
QY 1020 ENYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHSGHVFTRVKKKEVYM 1079
Db 1022 ENYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHSGHVFTRVKKKEVYM 1081
QY 1080 ALYNYLPGVFETVEMLPKSGAGIRWVECLIGEHLHAGMSTFLVYNSKQOTPLGMAAGHIR 1139
Db 1082 ALYNYLPGVFETVEMLPKSGAGIRWVECLIGEHLHAGMSTFLVYNSKQOTPLGMAAGHIR 1141
QY 1140 DFQITASGYGOWAPKARLHSGSINASTKEPFSWIKVDLAPMIHIGIKTOGARQKF 1199
Db 1142 DFQITASGYGOWAPKARLHSGSINASTKEPFSWIKVDLAPMIHIGIKTOGARQKF 1201
QY 1200 SSLYISQFTIMYSLDGKKWQTVYRGNSGTGLWFFGNDVSSGKKNIEPPIIARYLRHP 1259
Db 1202 SSLYISQFTIMYSLDGKKWQTVYRGNSGTGLWFFGNDVSSGKKNIEPPIIARYLRHP 1261

QY 1260 THYSIRSTLRMELMGCDLNSCMPLGMSKALSDAQITASSYFTNMFTMSPSKARLHLQ 1319
Db 1262 THYSIRSTLRMELMGCDLNSCMPLGMSKALSDAQITASSYFTNMFTMSPSKARLHLQ 1321
QY 1320 GRSNARWRQVNNPKEWLQVDFQTKMKVTGVTITQGVKSLTSMYKVEFLTSSSQDGHQWTL 1379
Db 1322 GRSNARWRQVNNPKEWLQVDFQTKMKVTGVTITQGVKSLTSMYKVEFLTSSSQDGHQWTL 1381
QY 1380 PFQNGKVKVFGQNGDSFTPVVNSLPPLLTRYLRHPSQSWHQAIALRMEVLGCEAQDLY 1438
Db 1382 PFQNGKVKVFGQNGDSFTPVVNSLPPLLTRYLRHPSQSWHQAIALRMEVLGCEAQDLY 1440

RESULT 8
AAE10832
ID AAE10832 standard; Protein; 1459 AA.
XX
AC AAE10832;
DT 18-DEC-2001 (first entry)
XX
DE Human factor VIII mutein encoded by vector pTGF8-2hyg-s.
KW Human; haemostatic; coagulant; blood clotting factor; factor VIII;
KW factor IX; therapy; haemophilia A; mutant; mutein.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /label= Signal_peptide
FT /product= "Human mature factor VIII mutein"
FT Region 760..775
FT /label= Linker-peptide
XX
PN WO200170968-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-EP03220.
XX
PR 22-MAR-2000; 2000EP-0106225.
PR 08-MAY-2000; 2000US-203249P.
XX
PA (OCTA-) OCTAGENE GMBH.
PI Hauser C, Hoerster A, Schroeder C, Lehnener M;
DR WPI: 2001-590175/66.
DR N-PSDB; AADI8176.
XX
PT Recombinantly producing human blood coagulation factors VIII and IX for
XX use in treating hemophilia -
PS Claim 17; Page 86-90; 104pp; English.
XX
CC The present invention relates to an improved method for the production
CC of recombinant human blood clotting factors, especially factors VIII and
CC IX, using an immortalised cell line stably expressing viral transcription
CC activators and carrying a vector comprising a promoter and a sequence
CC encoding the blood coagulation factor. The factor VIII mutein or a gene
CC transfer vector is used in the preparation of agents for treating
CC haemophilia, especially haemophilia A. The present sequence is
CC human factor VIII mutein encoded by vector pTGF8-2hyg-s. The vector
CC contains a silent mutation, resulting in a factor VIII mutein having
CC the substitution of the B-domain of wild-type factor VIII by the
CC linker peptide.
XX
SQ Sequence 1459 AA;

Query Match		99.2%; Score 7626; DB 22; Length 1459;	
Best Local Similarity		99.1%; Pred. No. 0;	
Matches 1431; Conservative		0; Mismatches 3; Indels 10; Gaps 2;	
QY	1	ATRRYVLGAVELSDWYQSDGLGELPVDARPPRPVPSPPFNTSVVYKTLRVFETVHLFN	60
DB	20	ATRRYVLGAVELSDWYQSDGLGELPVDARPPRPVPSPPFNTSVVYKTLRVFETVHLFN	79
QY	61	IAKPRPPMGLGPTTQAEYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ	120
DB	80	IAKPRPPMGLGPTTQAEYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ	139
QY	121	REKBDKVPFGSHTYVWVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR	180
DB	140	REKBDKVPFGSHTYVWVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR	199
QY	181	EGSLAKEKQTLHKFTLLFAVDEGKSWSETKNSLMQDRDAASARAWPKMTVNGYVNR	240
DB	200	EGSLAKEKQTLHKFTLLFAVDEGKSWSETKNSLMQDRDAASARAWPKMTVNGYVNR	259
QY	241	SLPGLIGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	300
DB	260	SLPGLIGCHRSYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	319
QY	301	MDLGOFLLCCHSSHQHDGMEAVKYVDSCEEPQLRWKNEAEYDDDLTDSMDVVERF	360
DB	320	MDLGOFLLCCHSSHQHDGMEAVKYVDSCEEPQLRWKNEAEYDDDLTDSMDVVERF	379
QY	361	DDNSPSFQIRSVAKKHPKTVHYIAABEEDWDYAPLVLPDDRYSKQYLNNGPQRTG	420
DB	380	DDNSPSFQIRSVAKKHPKTVHYIAABEEDWDYAPLVLPDDRYSKQYLNNGPQRTG	439
QY	421	RYKKVRPMAYTDETKTRAIOHESGILGLLYGVDGTLIIIFKXQASRPYNIYPHGI	480
DB	440	RYKKVRPMAYTDETKTRAIOHESGILGLLYGVDGTLIIIFKXQASRPYNIYPHGI	499
QY	481	TDVRPLYRRLPKGVKHLKDFPLPGEIFKIKWTVTVEDGPKSDPRCLTRYSSPVNME	540
DB	500	TDVRPLYRRLPKGVKHLKDFPLPGEIFKIKWTVTVEDGPKSDPRCLTRYSSPVNME	559
QY	541	RLASGLIGPLLCYKESVDQGNQIMSKRNVLFSVFDENRSWYLTENIQRFLEPNPAG	600
DB	560	RLASGLIGPLLCYKESVDQGNQIMSKRNVLFSVFDENRSWYLTENIQRFLEPNPAG	619
QY	601	VQLEDEPQASNMHSHNGVVDLSQVCLHEVAVWYILSGAQDFLSVFGSGYTFKH	660
DB	620	VQLEDEPQASNMHSHNGVVDLSQVCLHEVAVWYILSGAQDFLSVFGSGYTFKH	679
QY	661	KMYVEDTLTFPPSGETVFMSPENPGLWILGCHNSDFRNRGMTALLKVSCKDKNTGDYYE	720
DB	680	KMYVEDTLTFPPSGETVFMSPENPGLWILGCHNSDFRNRGMTALLKVSCKDKNTGDYYE	739
QY	721	DSYEDISAVLLSKXNAIEPRFSQNPVPLKHQR-----ETRRTYLSQDQBEIDYDDTI	774
DB	740	DSYEDISAVLLSKXNAIEPRFSQNPVPLKHQR-----SRHQAYRYRREGEITRTTTLQSDQBEIDYDDTI	795
QY	775	SVEMKEDFDIYDEDENQSPRSQKTRHYFIAAVERLMDYGVSSSPHVLNRAGSGSPV	834
DB	796	SVEMKEDFDIYDEDENQSPRSQKTRHYFIAAVERLMDYGVSSSPHVLNRAGSGSPV	855
QY	835	QFKXVVVFQBTGDSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSS	894
DB	856	QFKXVVVFQBTGDSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSS	915
QY	895	LISYEEDQAGAPRGNFKPNETKTYFWKQVHMAPTKDEPDKAWAYFSDVLEKDVH	954
DB	916	LISYEEDQAGAPRGNFKPNETKTYFWKQVHMAPTKDEPDKAWAYFSDVLEKDVH	975
QY	955	SGLIGPLLVCHTNTLNPAGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIOME	1014
DB	976	SGLIGPLLVCHTNTLNPAGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIOME	1035
QY	1015	DPTFKENYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKK	1074
DB	1036	DPTFKENYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKK	1095
QY	1075	BEYKKNALNLYPGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPLGWA	1134
DB	1096	BEYKKNALNLYPGVETVEMLPSPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPLGWA	1155
QY	1135	SGHIRDFQITASGQYGOVAPKRLARLHYSGSINAMSTKPPFWIKYVDLLAPMIHGIKTQG	1194
DB	1156	SGHIRDFQITASGQYGOVAPKRLARLHYSGSINAMSTKPPFWIKYVDLLAPMIHGIKTQG	1215
QY	1195	ARQKFSYLYISQFIIMYSLDGKKQTYRGNSTGTLMVFEFNVDSGSIKHNIENPPIIARY	1254
DB	1216	ARQKFSYLYISQFIIMYSLDGKKQTYRGNSTGTLMVFEFNVDSGSIKHNIENPPIIARY	1275
QY	1255	IRLHPHTYISIRSLRMELMGCDLNSCMPLGMSKAIISDAQITASSYFTNNFATWSPSKA	1314
DB	1276	IRLHPHTYISIRSLRMELMGCDLNSCMPLGMSKAIISDAQITASSYFTNNFATWSPSKA	1335
QY	1315	RLHLQGRSNAMRPQVNNPKWLQVDFQKTMKVTVTTQGVKSLLTSMYVKEFLISSSQDG	1374
DB	1336	RLHLQGRSNAMRPQVNNPKWLQVDFQKTMKVTVTTQGVKSLLTSMYVKEFLISSSQDG	1395
QY	1375	HWMTLFPQNGKVKVQGNQDSFTPVNSLDPLLTLYLRIHPQSWHQAIRMEVLGCEA	1434
DB	1396	HWMTLFPQNGKVKVQGNQDSFTPVNSLDPLLTLYLRIHPQSWHQAIRMEVLGCEA	1455
QY	1435	QDLY 1438	
DB	1456	QDLY 1459	
RESULT 9			
AAE10833			
ID	AAE10833	standard; Protein; 1459 AA.	
XX	AAE10833;		
AC	AAE10833;		
XX	18-DEC-2001	(first entry)	
DT	18-DEC-2001	(first entry)	
XX	Human factor VIII	mutin encoded by vector pTGF8-3.	
DE	Human factor VIII	mutin encoded by vector pTGF8-3.	
XX	Human; haemostatic; coagulant; blood clotting factor; factor VIII;		
KW	factor IX; therapy; haemophilia A; mutant; mutin.		
XX	Homo sapiens.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
FT	Peptide	1..19	
FT	Protein	/label= Signal_peptide	
FT	Misc-difference 181	/product= "Human mature factor VIII mutin"	
FT	Region	/note= "Wild type Val substituted with Ala"	
FT	Misc-difference 1350	/label= Linker-peptide	
FT	Misc-difference 1350	/note= "Wild type Val substituted with Glu"	
XX	WO200170968-A2.		
PN	27-SEP-2001.		
XX	21-MAR-2001; 2001WO-BP03220.		
PF	22-MAR-2000; 2000EP-0106225.		
XX	08-MAY-2000; 2000US-203249P.		
PR	(OCTA-) OCTAGENE GMBH.		
XX	Hauser C, Hoerster A, Schroeder C, Lehnner M;		

DR WPI; 2001-590175/66.
XX N-PSDB; AAD18176.
PT Recombinantly producing human blood coagulation factors VIII and IX for
XX use in treating hemophilia -
PS Claim 17; Page 99-103; 104pp; English.
XX
CC The present invention relates to an improved method for the production
CC of recombinant human blood clotting factors, especially factors VIII and
CC IX, using an immortalised cell line stably expressing viral transcription
CC activators and carrying a vector comprising a promoter and a sequence
CC encoding the blood coagulation factor. The factor VIII mutin or a gene
CC transfer vector is used in the preparation of agents for treating
CC haemophilia, especially haemophilia A. The present sequence is human
CC factor VIII mutin encoded by vector pTGF8-3. This sequence contains
CC 2 mutations and the B-domain of wild-type factor VIII is substituted
CC with a linker peptide.
XX
SQ Sequence 1459 AA;
Query Match 99.0%; Score 7616; DB 22; Length 1459;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1429; Conservative. 0; Mismatches 5; Indels 10; Gaps 2;
QY 1 ATRYYILGAVELSWDMQSDGLGELPVDARPPRPVPSPPNTSVVYKTLFVEFTVHLEN 60
DB 20 ATRYYILGAVELSWDMQSDGLGELPVDARPPRPVPSPPNTSVVYKTLFVEFTDHLN 79
QY 61 IAKPRPMMGLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKAEGAEYDDQTSQ 120
DB 80 IAKPRPMMGLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSYNKAEGAEYDDQTSQ 139
QY 121 REKEDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDDKVPFGSGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS LAKEKQTLLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASAPMPKMTVGVYNR 240
DB 200 EGS LAKEKQTLLHKFILLFAVFDGSKSWHSETKNSLMQDRDAASAPMPKMTVGVYNR 259
QY 241 SLRGLGCHRSVYWHVIGMTTPEVHSIFLEHTPLVNRHQASLEISPIFTLTAQTLL 300
DB 260 SLRGLGCHRSVYWHVIGMTTPEVHSIFLEHTPLVNRHQASLEISPIFTLTAQTLL 319
QY 301 MDLQGFLLFCHSSHQHDGMEAYVKVDSCEPFPQLMKMNEAEYDDDLTDSMDVYRF 360
DB 320 MDLQGFLLFCHSSHQHDGMEAYVKVDSCEPFPQLMKMNEAEYDDDLTDSMDVYRF 379
QY 361 DDNSPSFTQIRSVAKHPKTHVHTAAEEEDWDYAPLVAPDDRYSKSYQLNNGPQRIG 420
DB 380 DDNSPSFTQIRSVAKHPKTHVHTAAEEEDWDYAPLVAPDDRYSKSYQLNNGPQRIG 439
QY 421 RYKVKVRWAYTDETPKTRAEQHSGLGLPGLLYGRVGTLLIIPKQASRPVNYLPHGI 480
DB 440 RYKVKVRWAYTDETPKTRAEQHSGLGLPGLLYGEVGTLLIIPKQASRPVNYLPHGI 499
QY 481 TDVRLYSLRRLPKGVAKHLKDFPILFGEIFKYKWTVTVEDGPKSPRCITRYYSFVWNE 540
DB 500 TDVRLYSLRRLPKGVAKHLKDFPILFGEIFKYKWTVTVEDGPKSPRCITRYYSFVWNE 559
QY 541 RLASGLIGPLIICVKSVDQRGNQIMSDKRVNLLFSVPDENRWSYLTENIORFLPNPAG 600
DB 560 RLASGLIGPLIICVKSVDQRGNQIMSDKRVNLLFSVPDENRWSYLTENIORFLPNPAG 619
QY 601 VOLEPPEFOASINMISNGYVDSLSQLSVCLHEVAYWYLSIGATDFLSVFFSGVTKH 660
DB 620 VOLEPPEFOASINMISNGYVDSLSQLSVCLHEVAYWYLSIGATDFLSVFFSGVTKH 679
QY 661 KMVYEDTLTLFPFSGETVMSMENPGLMILGCHNSDFRNRGMTALLKVSCKDKNTGDYVE 720
DB 680 KMVYEDTLTLFPFSGETVMSMENPGLMILGCHNSDFRNRGMTALLKVSCKDKNTGDYVE 739

QY 721 DSYEDISAYLLSKNAIEPRSFSONPPVLKRQR-----EITRTTLOSDQBEIDYDDTI 774
DB 740 DSYEDISAYLLSKNAIEPRSFSON-----SRHQAYRYRGRGTRTTLQSDQBEIDYDDTI 795
QY 775 SYEMKKEPDFDIYDEDENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 834
DB 796 SYEMKKEPDFDIYDEDENQSPRSFQKTRHYFTAAVERLWDYGMSSSPHYLRNRAQSGSVP 855
QY 835 QPKKVVFQBPSTGSGSTOPLYRGELNEHGLIGPYIRAEVEDNIMWTPRQASRPSPYSS 894
DB 856 QPKKVVFQBPSTGSGSTOPLYRGELNEHGLIGPYIRAEVEDNIMWTPRQASRPSPYSS 915
QY 895 LLSYBEDORQAGPEPKNFVKPNETITYFKVQHHMPTKDFDCKAWAFSDVDLEKDVH 954
DB 916 LLSYBEDORQAGPEPKNFVKPNETITYFKVQHHMPTKDFDCKAWAFSDVDLEKDVH 975
QY 955 SGLIGPLVCHVNTNLPNPAHGRQVTVQEFALFTTIFDETKSWYFTENMENCRAPCNIQME 1014
DB 976 SGLIGPLVCHVNTNLPNPAHGRQVTVQEFALFTTIFDETKSWYFTENMENCRAPCNIQME 1035
QY 1015 DPTKENTRFRHAINGYIMDTLPGLVMAODORIRWYLLSNGSNENIHSIFSGHVTVRK 1074
DB 1036 DPTKENTRFRHAINGYIMDTLPGLVMAODORIRWYLLSNGSNENIHSIFSGHVTVRK 1095
QY 1075 BEYKMALYNLYPGVETVEMLPFSKAGIWRVECLIGEHLAGNSTLFLVYSNKCQPLGNA 1134
DB 1096 BEYKMALYNLYPGVETVEMLPFSKAGIWRVECLIGEHLAGNSTLFLVYSNKCQPLGNA 1155
QY 1135 SGHIRDFOITAGQYGOVAPKPLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGIKTG 1194
DB 1156 SGHIRDFOITAGQYGOVAPKPLARLHYSGSINAWSTKBPFSWIKVDLLAPMIHGIKTG 1215
QY 1195 ARQKFSYLSYFIIWYSLDGKKWQTYRGNSTGLMVRGNYDSSGIXHNIENPPIIARY 1254
DB 1216 ARQKFSYLSYFIIWYSLDGKKWQTYRGNSTGLMVRGNYDSSGIXHNIENPPIIARY 1275
QY 1255 IRLHPHTYSIRSLRMELMGCDLNSCMPLGMSKAIASDAQITASSYFTNMPATWSPSKA 1314
DB 1276 IRLHPHTYSIRSLRMELMGCDLNSCMPLGMSKAIASDAQITASSYFTNMPATWSPSKA 1335
QY 1315 RLHLAGRSNARPOVNNPKWLQVDFQKTMKVGTGTVTKVSKLLTSMVYKFLISSQDG 1374
DB 1336 RLHLAGRSNARPOVNNPKWLQVDFQKTMKVGTGTVTKVSKLLTSMVYKFLISSQDG 1395
QY 1375 HQWTLFFONGKVKVFGQNDQSTFPVNSIDPPLTRYLRHQPQSHWOIALRMEVLGCEA 1434
DB 1396 HQWTLFFONGKVKVFGQNDQSTFPVNSIDPPLTRYLRHQPQSHWOIALRMEVLGCEA 1455
QY 1435 QDLY 1438
DB 1456 QDLY 1459
RESULT 10
AAE10827
ID AAE10827 standard; Protein; 1459 AA.
XX AAE10827;
AC AAE10827;
XX AAE10827;
DT 18-DEC-2001 (first entry)
XX Human factor VIII mutin encoded by vector pTGF8-1.
DE Human; haemostatic; coagulant; blood clotting factor; factor VIII;
XX Human; therapy; haemophilia A; mutant; mutin.
KW factor IX; therapy; haemophilia A; mutant; mutin.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH 1..19
FT Peptide /label= Signal_peptide
FT Protein 20..1459

AAP80265

ID AAP80265 standard; protein; 1516 AA.

XX

AC AAP80265;

XX

DT 25-MAR-2003 (updated)

DT 10-OCT-1990 (first entry)

XX

DE Modified factor VIII:C sequence with the Q744-D1563 deletion.

XX

KW Modified factor VIII:C; maturation polypeptide; haemophilia;

KW

KW blood coagulation; QD deletion.

XX

OS Homo sapiens.

XX

PN WO8800831-A.

XX

PD 11-FEB-1988.

XX

PF 31-JUL-1987; 87WO-US01814.

XX

PR 01-AUG-1986; 86US-0893375.

XX

XX (BIOJ) BIOGEN NV.

PA

PA (PASE/) PASEK M P.

PI

PI Pasek MP;

XX

DR WPI; 1988-049866/07.

XX

DR N-PSDB; AAN80444.

XX

XX New DNA sequences encoding modified factor VIII:C - with deletion of DNA

PT

PT encoding maturation polypeptide, useful for high yield transformation.

XX

PS Claim 3; Page 51-52-53-54; 97pp; English.

XX

CC A major part of the sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1563. The QD deletion
 CC retains approximately 90 amino acids of the maturation polypeptide
 CC (four amino acids at the N-terminal end and 86 amino acids at
 CC the C-terminal end). The full length Factor VIII:C cDNA has two
 CC changes with respect to the published sequence (EPO application 160457):
 CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
 CC (Phe to Leu).

CC

CC The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced factor VIII:C and are more
 CC easily purified. The peptide is used for treating haemophilia A, both
 CC acute and prolonged bleeding.

CC

CC See also AAN80446 and AAN80447.

CC

CC (Updated on 25-MAR-2003 to correct PA field.)

XX

SQ Sequence 1516 AA;

Query Match

Best Local Similarity 99.0%; Score 7611.5; DB 9; Length 1516;

Matches 1434; Conservative 1; Mismatches 3; Indels 77; Gaps 2;

QY

1 ATRRYYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFNTSVVYKTLFVEFTVHLFN 60

Db

2 ATRRYYLGAVELSWDMQSDLGELPVDARFPFRVPKSPFNTSVVYKTLFVEFTVHLFN 61

QY

61 IAKPRPPNGLLGPPIQAEVYDVTVITLKNMASHVPSLHGVSVTWKASEGAEYDDQTSQ 120

Db

62 IAKPRPPNGLLGPPIQAEVYDVTVITLKNMASHVPSLHGVSVTWKASEGAEYDDQTSQ 121

QY

121 REKEDKVPFGSGSHYVQVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVCR 180

Db

122 REKEDKVPFGSGSHYVQVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVCR 181

QY

181 EGSLAKTKQTLLHKFILLFAVFDEKSWHSETKNSLMQDRDAASARAWPKQHTVNGYVNR 240

Db

182 EGSLAKTKHTLHKFILLFAVFDEKSWHSETKNSLMQDRDAASARAWPKQHTVNGYVNR 241

241 S-LPGLIGCHRSKSVVHWVGTTPEVHSIFLEIGHTFLVRNHRQASLEISPTITLTAQTL 299
 242 SLPLGLIGCHRSKSVVHWVGTTPEVHSIFLEIGHTFLVRNHRQASLEISPTITLTAQTL 301
 300 LMDLQGLFLFCHISSHQHDMGMEAYVKVDSCPEEPQLRMKNNEAEADYDDDLTDSBMDVVR 359
 302 LMDLQGLFLFCHISSHQHDMGMEAYVKVDSCPEEPQLRMKNNEAEADYDDDLTDSBMDLVR 361
 360 EDDNDSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDRSKYSQVNLNNGPQRI 419
 362 EDDNDSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLVLPADDRSKYSQVNLNNGPQRI 421
 420 GRKYKKVRFMAYTDETFKTRAEIOHESGILGLLAGEVGDILLIIFKQASRPYNYIPHG 479
 422 GRKYKKVRFMAYTDETFKTRAEIOHESGILGLLAGEVGDILLIIFKQASRPYNYIPHG 481
 480 ITDVRPLYSRRLPKGVKHLKDPPIIPGIFIKYKNTVTVEDGPTKSDPRCLTRYYSFVNM 539
 482 ITDVRPLYSRRLPKGVKHLKDPPIIPGIFIKYKNTVTVEDGPTKSDPRCLTRYYSFVNM 541
 540 EBDLASGLIGPLLCYKESVDQGNQIMSDKNVILFSVFDNRSWYLTENIQRFLPNPA 599
 542 EBDLASGLIGPLLCYKESVDQGNQIMSDKNVILFSVFDNRSWYLTENIQRFLPNPA 601
 600 GVQLEDPEFOASNIMHSINGVYVDSLQLSVCLHEVAYVYILSIGAQDTFLSVFFSGYTFK 659
 602 GVQLEDPEFOASNIMHSINGVYVDSLQLSVCLHEVAYVYILSIGAQDTFLSVFFSGYTFK 661
 660 HKMVTEDTLTLPFSGSETVFMSENPGLWILGCHNSDFNRGMTALLKVSCKNTGDYV 719
 662 HKMVTEDTLTLPFSGSETVFMSENPGLWILGCHNSDFNRGMTALLKVSCKNTGDYV 721
 720 EDSYEDISAYLLSKNNAIEPRSPF----- 742
 722 EDSYEDISAYLLSKNNAIEPRSPF----- 742
 743 -----SONPVLKXHQREITRTLLQS 763
 762 ILSLNACSNHAIANAINGQNKPELEVWAKQGRTERLCSQNPVVLKXHQREITRTLLQS 841
 764 DQEEIDYDITTSVEMKEDDFDIYDEDENQSPRSFQKTRHYPIAAVERLWDYGMSSSPHV 823
 842 DQEEIDYDITTSVEMKEDDFDIYDEDENQSPRSFQKTRHYPIAAVERLWDYGMSSSPHV 901
 824 LRNRQAQSGSVQFKKVVQFQFTDQSGFTQPLRGELNEHLGLIGYIRAEVEDNIMVTFRN 883
 902 LRNRQAQSGSVQFKKVVQFQFTDQSGFTQPLRGELNEHLGLIGYIRAEVEDNIMVTFRN 961
 884 QASRPYSFYSSLSIYSEEDQRCQABPRKNFVKPNETKTYFWKVQHHMAPTKDBFDCWAY 943
 962 QASRPYSFYSSLSIYSEEDQRCQABPRKNFVKPNETKTYFWKVQHHMAPTKDBFDCWAY 1021
 944 FSDVDLEKDVHSLIGLPLLVCHTTLNPAHQRVTVQEFALFFPTIFDTSKSWYFENMER 1003
 1022 FSDVDLEKDVHSLIGLPLLVCHTTLNPAHQRVTVQEFALFFPTIFDTSKSWYFENMER 1081
 1004 NCRAPCNQIMEDPTFKENYREHANGYIMDTLPGVMAQODRIEYVILLSMGNSNENIHSIH 1063
 1082 NCRAPCNQIMEDPTFKENYREHANGYIMDTLPGVMAQODRIEYVILLSMGNSNENIHSIH 1141
 1064 FSGHVFTRKKEEYKMAVLYNLYPGVFFTEVMLPSKAGIWRVECLIGELHAGMSTFLVY 1123
 1142 FSGHVFTRKKEEYKMAVLYNLYPGVFFTEVMLPSKAGIWRVECLIGELHAGMSTFLVY 1201
 1124 SNKQTPILGASGHIRDFQITAGSQYQWAPKARLHYSGINAWSTKEPFSWIKVDLLA 1183
 1202 SNKQTPILGASGHIRDFQITAGSQYQWAPKARLHYSGINAWSTKEPFSWIKVDLLA 1261
 1184 PMIHGKTKQARCKESSLYISOFIMYSLDGKKQWYRGNSGTGLMVFGNVDSSGIKH 1243
 1262 PMIHGKTKQARCKESSLYISOFIMYSLDGKKQWYRGNSGTGLMVFGNVDSSGIKH 1321
 1244 NIFNPPIIARYIRLHPHYISIRSTLRMLMGCDLNSCSMPLGMSKASDAQITASSYFT 1303

Db 1322 NIFNPPITARYIRLHETHYSIRSTRKMLMGDLNCSMPLGMSKASDAQITASSYFT 1381
QY 1304 NMFATWSPSKARLHLQGRSNARWPQVNNPKELQVDFQTKMKVGTGTTQGVKSLTSMYV 1363
Db 1382 NMFATWSPSKARLHLQGRSNARWPQVNNPKELQVDFQTKMKVGTGTTQGVKSLTSMYV 1441
QY 1364 KEFLISSODGHQWTLFPQNGKVKVQGNQDSFTPVNSLDPPLTRYLRIRIIPQSNVHQI 1423
Db 1442 KEFLISSODGHQWTLFPQNGKVKVQGNQDSFTPVNSLDPPLTRYLRIRIIPQSNVHQI 1501
QY 1424 ALRMEVLGCEAODLY 1438
Db 1502 ALRMEVLGCEAODLY 1516

RESULT 12

AAB48842
ID AAB48842 standard; protein; 1424 AA.

XX
AC AAB48842;

XX 13-MAR-2001 (first entry)

XX Mutant mature human factor VIII, SEQ ID NO:5.

XX Factor VIII; human; B domain; LRP-mediated plasma clearance;
KW receptor-dependent clearance; receptor-independent clearance;
KW half-life; haemophilia; mutant; muten.

XX Homo sapiens.

XX WO200071714-A2.

XX 30-NOV-2000.

XX 24-MAY-2000; 2000WO-US14111.

XX 24-MAY-1999; 99US-0135847.

XX (AMNA-) AMERICAN NAT RED CROSS.

XX Saenko EL, Strickland DK;

XX WPI; 2001-025163/03.

XX Factor VIII mutants having increased half-life useful for treating
PT hemophilia, comprise one or more amino acid substitutions in the A2
PT and/or C2 domain of factor VIII -

XX Claim 9; Fig 2A-B; 121pp; English.

XX The invention relates to human factor VIII mutants comprising an amino
CC acid substitution at one or more positions in the A2 domain and/or an
CC amino acid substitution at one or more positions in the C2 domain.
CC The invention also encompasses a factor VIII mutant which lacks a B
CC domain (AAB48842). The factor VIII mutants have an increased half-life
CC in the bloodstream. The A2 domain mutants exhibit reduced LRP-dependent
CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants
CC have reduced receptor-independent clearance. The invention also relates
CC to a method of using RAP (receptor associated protein), a protein which
CC inhibits LRP (low density lipoprotein related protein)-mediated ligand
CC internalisation, to increase the half-life of factor VIII. The mutant
CC factor VIII proteins, and nucleotides encoding them, are useful
CC for treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and
CC nucleic acids encoding them may also be used in the treatment of
CC haemophilia. In combination with a mutant factor VIII protein or DNA of
CC the invention. The invention provides means of increasing the half-life
CC of factor VIII by reducing its clearance from plasma. The present
CC sequence represents a mutant mature human factor VIII which lacks a B
CC domain.

XX Sequence 1424 AA;

SQ

Query Match		98.7%;	Score 7592;	DB 22;	Length 1424;
Best Local Similarity		99.0%;	Pred. No. 0;		
Matches 1423;		Conservative	0;	Mismatches	1; Indels 14; Gaps 1;
QY	1	ATRRYYLGAVELSDWYQSDIGELPVDARPPRPVPSPPNTSVVYKKTIVFEFTVHLFN	60		
Db	1	ATRRYYLGAVELSDWYQSDIGELPVDARPPRPVPSPPNTSVVYKKTIVFEFTVHLFN	60		
QY	61	IAKPRPPMGLLGPTIQAEVYDVTVVITLKMAASHPVSLHVGVSYWKASEGAEYDDQTSQ	120		
Db	61	IAKPRPPMGLLGPTIQAEVYDVTVVITLKMAASHPVSLHVGVSYWKASEGAEYDDQTSQ	120		
QY	121	REKEDDKVPFGSHTYVWQVLEKNGPMASDPLCLTYSYLSHVLDVLUINSGLIGALLVCR	180		
Db	121	REKEDDKVPFGSHTYVWQVLEKNGPMASDPLCLTYSYLSHVLDVLUINSGLIGALLVCR	180		
QY	181	EGSLAKEKQTQTLHKFILLFAVPDEGKSWHSTKSLMODRDAASARAWPKWHTVNGYNR	240		
Db	181	EGSLAKEKQTQTLHKFILLFAVPDEGKSWHSTKSLMODRDAASARAWPKWHTVNGYNR	240		
QY	241	SLPGLIGCHRSVYVHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL	300		
Db	241	SLPGLIGCHRSVYVHVIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQLL	300		
QY	301	MDLQGFLLFCHISSHQHGMAYVKVDSCEPQLRMKQNEEAEDYDDDLTSEMDVVRP	360		
Db	301	MDLQGFLLFCHISSHQHGMAYVKVDSCEPQLRMKQNEEAEDYDDDLTSEMDVVRP	360		
QY	361	DDNSPSFIQIERSVAKGHPKTVVHVIABEEDWDYAPLVLAPDDRSYKQVNLNNGPORG	420		
Db	361	DDNSPSFIQIERSVAKGHPKTVVHVIABEEDWDYAPLVLAPDDRSYKQVNLNNGPORG	420		
QY	421	RKYKVRFMAYTDEFTKREAIQHESGILGPLLXGEVGDITLLIFKNQASRYNIYPHGI	480		
Db	421	RKYKVRFMAYTDEFTKREAIQHESGILGPLLXGEVGDITLLIFKNQASRYNIYPHGI	480		
QY	481	TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTGKSDPRCLTYYSSFVNM	540		
Db	481	TDVRPLYSRRLPKGVKHLKDPFLLPGEIFKYKWTVTVEDGTGKSDPRCLTYYSSFVNM	540		
QY	541	RDLASGLIGPLLCYKESVDORGQIMSDKRNVLFSVFDENRSHYLTENIQRFLENPAG	600		
Db	541	RDLASGLIGPLLCYKESVDORGQIMSDKRNVLFSVFDENRSHYLTENIQRFLENPAG	600		
QY	601	VLEDDPPEQASNMHSINGYVDFSQLSVCLHEVAYWYLSIGACTDPLSFYFGYTFKH	660		
Db	601	VLEDDPPEQASNMHSINGYVDFSQLSVCLHEVAYWYLSIGACTDPLSFYFGYTFKH	660		
QY	661	KWYVEDTLTLPFSGETVFMSENPGWLILGCHNSDRNRGWTALLKVSSCKNTGDYIE	720		
Db	661	KWYVEDTLTLPFSGETVFMSENPGWLILGCHNSDRNRGWTALLKVSSCKNTGDYIE	720		
QY	721	DSYEDISAYLLSKNNAIEPRFSQNPVYLKREHREITRTTLQSDQEEIDYDDTISVEMKK	780		
Db	721	DSYEDISAYLLSKNNAIEPRFSQNPVYLKREHREITRTTLQSDQEEIDYDDTISVEMKK	780		
QY	781	EDFIYDEDENOQSPRQKTRHYFTAAVERLWYDGMSSSPHVLNRQAQSGSVQFKKV	840		
Db	781	EDFIYDEDENOQSPRQKTRHYFTAAVERLWYDGMSSSPHVLNRQAQSGSVQFKKV	840		
QY	841	FOEFTDGSFTQPLVRGELNEHLGLLGPVTRAIVEDNTMTVFNQASRPYFSYSLISVEE	900		
Db	841	FOEFTDGSFTQPLVRGELNEHLGLLGPVTRAIVEDNTMTVFNQASRPYFSYSLISVEE	900		
QY	901	DQRQASPRKNFVAPNETKTYFMKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLTGP	960		
Db	901	DQRQASPRKNFVAPNETKTYFMKVQHMAPTKDEFCCKAWAYFSDVLEKDVHSGLTGP	960		
QY	961	LLVCHTNTLPAHGRQVTVQEFALFFTYFDRTKSWYFTENMERNCRAPCNIMQEDPTFKE	1020		
Db	961	LLVCHTNTLPAHGRQVTVQEFALFFTYFDRTKSWYFTENMERNCRAPCNIMQEDPTFKE	1020		

QY 1021 NYRFAHNGYIMDTLPGLVMAQQRIRWYLLSGNSNENIHSIHFSGHVFTVRKKEBKMA 1080
 DB |||||
 QY 1007 NYRFAHNGYIMDTLPGLVMAQQRIRWYLLSGNSNENIHSIHFSGHVFTVRKKEBKMA 1066
 DB |||||
 QY 1081 LYNLYPGVFTVEMLPSKAGIWRVECLIGHHAGHASTLPVYSNKCQPTPLGASGHIRD 1140
 DB |||||
 QY 1067 LYNLYPGVFTVEMLPSKAGIWRVECLIGHHAGHASTLPVYSNKCQPTPLGASGHIRD 1126
 DB |||||
 QY 1141 FQITASQYQWAPKLARLHYSGSINAWSTKEPFWIKVDLLAPMIIHGIKTOGAROKFS 1200
 DB |||||
 QY 1127 FQITASQYQWAPKLARLHYSGSINAWSTKEPFWIKVDLLAPMIIHGIKTOGAROKFS 1186
 DB |||||
 QY 1201 SLVYSQFIIMYSIDGKQWYRGNSGTGLMVFNGVDSSGIKHIFNPPIIARILHPT 1260
 DB |||||
 QY 1187 SLVYSQFIIMYSIDGKQWYRGNSGTGLMVFNGVDSSGIKHIFNPPIIARILHPT 1246
 DB |||||
 QY 1261 HYSIRSTLRMLGCDLNSCSPLGMSKASDAQITASSYFTNMFTATSPSKARLHLQ 1320
 DB |||||
 QY 1247 HYSIRSTLRMLGCDLNSCSPLGMSKASDAQITASSYFTNMFTATSPSKARLHLQ 1306
 DB |||||
 QY 1321 RSNAMRPQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFLISSODGHOWTLF 1380
 DB |||||
 QY 1307 RSNAMRPQVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFLISSODGHOWTLF 1366
 DB |||||
 QY 1381 FONGKVKVFGNQDSTPTPVNSLDPLLRILYLRHFSQWVHQIALRNEVLGCEAQDLY 1438
 DB |||||
 QY 1367 FONGKVKVFGNQDSTPTPVNSLDPLLRILYLRHFSQWVHQIALRNEVLGCEAQDLY 1424
 DB |||||

RESULT 13

AAO18622 standard; Protein; 1424 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

CC reduced sulfate proteoglycan (HSPG)-dependent or receptor-independent
 CC clearance and procoagulant activity, which has a nonconservative amino
 CC acid substitution at one or more positions in the A2 domain consisting of
 CC Lys(380, 512, 556, 570 or 659) or Arg(490, 527, 562 or 571) or in the C2
 CC domain relative to the wild-type. The mutant factor VIII or the
 CC polynucleotide encoding it and a receptor-associated protein (alpha2
 CC macroglobulin receptor-associated protein or RAP) are useful for treating
 CC haemophilia. The mutated protein has a longer half-life. The present
 CC sequence is the human mature B-domainless factor VIII protein.
 XX
 SQ Sequence 1424 AA;

Query Match 98.7%; Score 7592; DB 23; Length 1424;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1423; Conservative 0; Mismatches 1; Indels 14; Gaps 1;

QY 1 ATRRYVLGAVELSWDMQSDLGELPVDARFPVRPKSPFPNTSVVYKTLFVEFTVHLFN 60
 DB |||||
 QY 61 IAKPRPPWMLGPTIOAEVDTVVITLKMAASHPVSLHGVSVWKASEGAYDDQTSQ 120
 DB |||||
 QY 61 IAKPRPPWMLGPTIOAEVDTVVITLKMAASHPVSLHGVSVWKASEGAYDDQTSQ 120
 DB |||||
 QY 121 REKEDDKVFPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 DB |||||
 QY 121 REKEDDKVFPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 DB |||||
 QY 181 EGSLAKETQTLHKPILLFAVDFDEKSWHSETKNSLMQDRDAASARAWPKMHTVGVNR 240
 DB |||||
 QY 181 EGSLAKETQTLHKPILLFAVDFDEKSWHSETKNSLMQDRDAASARAWPKMHTVGVNR 240
 DB |||||
 QY 241 SLPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTPLVRNHRQASLEISPIFLTAQTLL 300
 DB |||||
 QY 241 SLPLGLIGCHRSVYWHVIGMGTTPVHSIFLEGHTPLVRNHRQASLEISPIFLTAQTLL 300
 DB |||||
 QY 301 MDLQGLFLFCHISSHQHDGMEAYVKVDSCEPQOLRMKNNEAEYDDDLTDSMDVVRP 360
 DB |||||
 QY 301 MDLQGLFLFCHISSHQHDGMEAYVKVDSCEPQOLRMKNNEAEYDDDLTDSMDVVRP 360
 DB |||||
 QY 361 DDNSPSFTQIRSVAKKHPKTVVHYIAAEEEDWDYAPLVAPDDRYSYQSLNNGPQRIG 420
 DB |||||
 QY 361 DDNSPSFTQIRSVAKKHPKTVVHYIAAEEEDWDYAPLVAPDDRYSYQSLNNGPQRIG 420
 DB |||||
 QY 421 RYKVKVREWAYTDETFKTRTAEAOHESGILGPLLYGEVGTLLIIFKNQASRPNTYPHGI 480
 DB |||||
 QY 421 RYKVKVREWAYTDETFKTRTAEAOHESGILGPLLYGEVGTLLIIFKNQASRPNTYPHGI 480
 DB |||||
 QY 481 TDVRLYSRRLLPKGVKHLKDFPILPGEIFKVKWTVTVEDGPTKSDPRCLTRYYSFVWME 540
 DB |||||
 QY 481 TDVRLYSRRLLPKGVKHLKDFPILPGEIFKVKWTVTVEDGPTKSDPRCLTRYYSFVWME 540
 DB |||||
 QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLNPAG 600
 DB |||||
 QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLNPAG 600
 DB |||||
 QY 601 VQLEDPEFQASNIMHSINGVYVDSIQLSVCLHEVAYVYILSIGACTDPLSFYFSGYTKH 660
 DB |||||
 QY 601 VQLEDPEFQASNIMHSINGVYVDSIQLSVCLHEVAYVYILSIGACTDPLSFYFSGYTKH 660
 DB |||||
 QY 661 KQVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFNRGMTALLKVSSCDKNTGDYVE 720
 DB |||||
 QY 661 KQVYEDTLTLFPFSGETVFMENPGLWILGCHNSDFNRGMTALLKVSSCDKNTGDYVE 720
 DB |||||
 QY 721 DSYEDISAYLLSKNAIEPRSPQNPVLPKHQREITRTTLQSQDEEIDYDDTISVENMK 780
 DB |||||
 QY 721 DSYEDISAYLLSKNAIEPRSPQNPVLPKHQREITRTTLQSQDEEIDYDDTISVENMK 780
 DB |||||
 QY 781 EDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQPKKVV 840
 DB |||||
 QY 781 EDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGSVQPKKVV 840
 DB |||||
 QY 841 FQETDGSFTQPLRYRGLNHLGLLGPYIRAEVEDNINVTFRNQASRPYSPYSLISYEE 900
 DB |||||

New mutant factor VIII with reduced sulfate proteoglycan
 (HSPG)-dependent or receptor-independent clearance and procoagulant
 activity for treating hemophilia -

Claim 9; Fig 12; 161pp; English.

The present invention relates to a mutant factor VIII protein with


```

Db      827 FQSFDTGSFTQPLVRGELNEHLGLLGYIIRAEVEDNIMVTRFQASRPYSFYSLISYEE 886
Qy      901 DQSGAEPRKFNFKPNETKTYFMKVQHMAPTDEDFCKAWAYFSDVDLEKDVHSGLIGP 960
Db      887 DQSGAEPRKFNFKPNETKTYFMKVQHMAPTDEDFCKAWAYFSDVDLEKDVHSGLIGP 946
Qy      961 LLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNQIOMEDPTTKE 1020
Db      947 LLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNQIOMEDPTTKE 1006
Qy      1021 NYRPHALNGYIMDTPLGLVNAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYMA 1080
Db      1007 NYRPHALNGYIMDTPLGLVNAQDQIRWYLLSGNSNENIHSIHFSGHVFTVRKKEEYMA 1066
Qy      1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQTPGLMASGHIRD 1140
Db      1067 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQTPGLMASGHIRD 1126
Qy      1141 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQARQKFS 1200
Db      1127 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQARQKFS 1186
Qy      1201 SLVYSQFIIMYSLDGKKWQYRGNSTGTLVFFGNVDSSGKINIFNPPIIARYIRLHPT 1260
Db      1187 SLVYSQFIIMYSLDGKKWQYRGNSTGTLVFFGNVDSSGKINIFNPPIIARYIRLHPT 1246
Qy      1261 HYSIRSTLRMLGCDLNSCMLPGMESKAISDAQITASSYFTNMFATWSPPSKARLHLOG 1320
Db      1247 HYSIRSTLRMLGCDLNSCMLPGMESKAISDAQITASSYFTNMFATWSPPSKARLHLOG 1306
Qy      1321 RSNARPOVNNPKWLOVDQKTKVGTGVTQGVKSLLTSMYKKEFLISSQDGHQWTLF 1380
Db      1307 RSNARPOVNNPKWLOVDQKTKVGTGVTQGVKSLLTSMYKKEFLISSQDGHQWTLF 1366
Qy      1381 FQNGKVKVFGQNDQSFPPVNSLDPPLLTTRYLRHPQSWVHQIALRMEVLGCEAQDLY 1438
Db      1367 FQNGKVKVFGQNDQSFPPVNSLDPPLLTTRYLRHPQSWVHQIALRMEVLGCEAQDLY 1424

```

RESULT 14

ABG92541 standard; Protein; 1447 AA.

XX AC ABG92541;

XX DT 19-NOV-2002 (first entry)

XX DE 5Arg B-domain-deleted-Factor VIII (FVIII).

XX KW Human; alpha-galactosidase; Factor VIII; Factor IX; gene therapy.

XX OS Homo sapiens.

XX OS Synthetic.

XX XX WO200264799-A2.

XX XX 22-AUG-2002.

XX XX 11-OCT-2001; 2001WO-US42655.

XX XX 11-OCT-2000; 2000US-0686497.

XX XX (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX XX Seldon RF, Miller AM, Treco DS;

XX XX MP1; 2002-627600/67.

XX XX N-PSDB; ABS68000.

XX PT New synthetic nucleic acid sequence involving a continuous stretch of a
 PT least 150 common codons, useful for expressing mammalian, preferably

human proteins e.g. alpha-galactosidase or Factor VIII or IX or for gene therapy

Example 1; Figure 9; 115pp; English.

The invention relates to a synthetic nucleic acid sequence, where at least one non-common or less-common codon is replaced with a common codon. The synthetic nucleic acid is useful for expressing mammalian, preferably human proteins e.g. alpha-galactosidase or Factor VIII or IX or for gene therapy. The synthetic nucleic acid allows precise dosing and reduces treatment costs. It is simple to apply in treating patients and is curative (one gene therapy treatment has the potential to last a patient's lifetime. ABG92540-ABG92541 represent Factor VIII and Factor IX synthetic amino acid sequences of the invention.

XX Sequence 1447 AA;

Query Match 98.7%; Score 7588; DB 23; Length 1447;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1423; Conservative 1; Mismatches 4; Indels 10; Gaps 1;

```

Qy      1 ATERYYLGAVELSWDMYQSDLGELPVDARFPVRPKSPFPNTSVVYKTLFVEFTVHLEN 60
Db      20 ATERYYLGAVELSWDMYQSDLGELPVDARFPVRPKSPFPNTSVVYKTLFVEFTVHLEN 79
Qy      61 IAKPRPPMGLLGPTIOAEVYDTVTTLQMAASHPVSLHAGVSVYKASGEGAYDDQTSQ 120
Db      80 IAKPRPPMGLLGPTIOAEVYDTVTTLQMAASHPVSLHAGVSVYKASGEGAYDDQTSQ 139
Qy      121 REKEDDKVPFGSGHTVVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
Db      140 REKEDDKVPFGSGHTVVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 199
Qy      181 EGS�AKEKQTTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYVNR 240
Db      200 EGS�AKEKQTTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVANGYVNR 259
Qy      241 SLFGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTLL 300
Db      260 SLFGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTLL 319
Qy      301 MDLQGFLLFCHISSHQHDGMEAYVKVDCSCPEPQQLRMKNRBAEDYDDDLTDTSEMDVVR 360
Db      320 MDLQGFLLFCHISSHQHDGMEAYVKVDCSCPEPQQLRMKNRBAEDYDDDLTDTSEMDVVR 379
Qy      361 DDNSPSFTQIRSVAKKQKTKWVHYIAABEEDWDYAPLVLAAPDRSYKSQYLNNGPQRIG 420
Db      380 DDNSPSFTQIRSVAKKQKTKWVHYIAABEEDWDYAPLVLAAPDRSYKSQYLNNGPQRIG 439
Qy      421 RKVKKVRFMAYTDETFKTREAIQHSGLIPGLLYGEVGDTLIIIFKNQASRYNLYPHGI 480
Db      440 RKVKKVRFMAYTDETFKTREAIQHSGLIPGLLYGEVGDTLIIIFKNQASRYNLYPHGI 499
Qy      481 TDVRELYSRRLPKGVKHLKDFPILFGEIFKYKWTVTVEDGPKTSPRCITRYYSFVNME 540
Db      500 TDVRELYSRRLPKGVKHLKDFPILFGEIFKYKWTVTVEDGPKTSPRCITRYYSFVNME 559
Qy      541 RDLASGLIGPLLI CYKESVDQGNQINSDKRNVLILFSVFDENRSHYLTENIQRFLENPAG 600
Db      560 RDLASGLIGPLLI CYKESVDQGNQINSDKRNVLILFSVFDENRSHYLTENIQRFLENPAG 619
Qy      601 VQLEDPFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGACTDPLSVFFSGYTPKH 660
Db      620 VQLEDPFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGACTDPLSVFFSGYTPKH 679
Qy      661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNRGWTALLKVSSCDKXTGDYYE 720
Db      680 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNRGWTALLKVSSCDKXTGDYYE 739
Qy      721 DSYEDISAYLLSKNNAIEPRFSQNPVPLKQHQREITRTTTLQSDQEEIDYDDTISVENMK 780
Db      740 DSYEDISAYLLSKNNAIEPRFSQNPVPLKQHQREITRTTTLQSDQEEIDYDDTISVENMK 789

```

781 EDYDIYDEENQSPRSFQKTRHYFIAAVERLWDYCMSSSPHVRNRAQSGVPOPKVY 840
 DB |||||
 790 EDYDIYDEENQSPRSFQKTRHYFIAAVERLWDYCMSSSPHVRNRAQSGVPOPKVY 849
 QY |||||
 DB |||||
 841 FQFTDGSFTQPLRYGELNEHGLGPGYIAEVEDNIMVTFRNQASRPYFYSLSIYEE 900
 DB |||||
 850 FQFTDGSFTQPLRYGELNEHGLGPGYIAEVEDNIMVTFRNQASRPYFYSLSIYEE 909
 QY |||||
 DB |||||
 901 DORQGAEPKFNFKPNETKTYFMKVQHMAPTKDEDFCKAWAFSDVDLEKDVHSLGIP 960
 DB |||||
 910 DORQGAEPKFNFKPNETKTYFMKVQHMAPTKDEDFCKAWAFSDVDLEKDVHSLGIP 969
 QY |||||
 DB |||||
 961 LNVCHTNLNPANGRQVTVQEFALFTTFIDETKSWYFTENMERNCRAPCNIMQEDPTFXE 1020
 DB |||||
 970 LNVCHTNLNPANGRQVTVQEFALFTTFIDETKSWYFTENMERNCRAPCNIMQEDPTFXE 1029
 QY |||||
 DB |||||
 1021 NYRFHAINGYIMDTLPLGLVMAQDQRTWVLLSGNSNENIHSIFSGHVTYVKKEYYKVA 1080
 DB |||||
 1030 NYRFHAINGYIMDTLPLGLVMAQDQRTWVLLSGNSNENIHSIFSGHVTYVKKEYYKVA 1089
 QY |||||
 DB |||||
 1081 LYNLYPGVETVEMLPKSKAGIWRVECLIGEHLAGNSTLFLVYSNKCQTPLGWASGHIRD 1140
 DB |||||
 1090 LYNLYPGVETVEMLPKSKAGIWRVECLIGEHLAGNSTLFLVYSNKCQTPLGWASGHIRD 1149
 QY |||||
 DB |||||
 1141 FQITASGOYQWAPKLARLHYSGSINAWTKPFSWIKVDLLAPMIHGIKTQGAQPKFS 1200
 DB |||||
 1150 FQITASGOYQWAPKLARLHYSGSINAWTKPFSWIKVDLLAPMIHGIKTQGAQPKFS 1209
 QY |||||
 DB |||||
 1201 SLVYSOFTIWSLDGKKWOTYRGNSTGLMVFPGNVDSGSIKHNFNPPIIARYRLHPT 1260
 DB |||||
 1210 SLVYSOFTIWSLDGKKWOTYRGNSTGLMVFPGNVDSGSIKHNFNPPIIARYRLHPT 1269
 QY |||||
 DB |||||
 1261 HYSIRSTLRMELMGCDLNSCMPLGMSKAI SDAQITASSYFTNMFATWSPSKARLHQQ 1320
 DB |||||
 1270 HYSIRSTLRMELMGCDLNSCMPLGMSKAI SDAQITASSYFTNMFATWSPSKARLHQQ 1329
 QY |||||
 DB |||||
 1321 RSNAMPQVNNPKWLVQDFQTKMKTGTTQGVKSLLTSMYVKFPLISSQDGHQWTLF 1380
 DB |||||
 1330 RSNAMPQVNNPKWLVQDFQTKMKTGTTQGVKSLLTSMYVKFPLISSQDGHQWTLF 1389
 QY |||||
 DB |||||
 1381 FONGKVKVFGQNDSFTPVNSLDPLLRILRIHPQSVVHGIARMEVLGCEAODLY 1438
 DB |||||
 1390 FONGKVKVFGQNDSFTPVNSLDPLLRILRIHPQSVVHGIARMEVLGCEAODLY 1447

RESULT 15
 AAP80267

ID AAP80267 standard; protein; 1425 AA.

AC AAP80267;

XX 25-MAR-2003 (updated)

DT 10-OCT-1990 (first entry)

DE Modified factor VIII:C sequence with the R740-E1649 deletion.

DE Modified factor VIII:C; maturation polypeptide; haemophilia;

DE blood coagulation; RE deletion.

OS Homo sapiens.

PN WO8800831-A.

XX 11-FEB-1988.

XX 31-JUL-1987; 87WO-US01814.

XX 01-AUG-1986; 86US-0893375.

XX (BIOJ) BIOGEN NV.

PA (PASE/) PASEK M P.

XX Pasek MP;

XX WPI: 1988-049866/07.
 DR N-P8DB; AAN80446.
 XX New DNA sequences encoding modified factor VIII:C - with deletion of DNA
 PT encoding maturation polypeptide, useful for high yield transformation.
 XX Claim 3; Page 57-58-59-60; 97pp; English.
 XX The entire sequence encoding the maturation polypeptide of
 CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
 CC The full length factor VIII:C cDNA has two changes with respect to the
 CC published sequence (BPO application 160457):
 CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
 CC (Phe to Leu). The product is produced in approx. 20 times higher
 CC yields than previous recombinant produced in approx. 20 times higher
 CC acute and prolonged bleeding.
 CC See also AAN80444 and AAN80447.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX Sequence 1425 AA;

Query Match 98.6%; Score 7581; DB 9; Length 1425;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1421; Conservative 0; Mismatches 3; Indels 14; Gaps 1;

QY 1 ATRRYYLGAVELSWNDYMQSDIGELPVDARFPFRVPKSPFNSTVYVKTLFVFEVTHLPN 60
 DB |||||
 2 ATRRYYLGAVELSWNDYMQSDIGELPVDARFPFRVPKSPFNSTVYVKTLFVFEVTHLPN 61
 QY 61 IAKPPPPMGLLGPTIOAEVYDVTVTILKNMASHVSLHAGVSVYKASEGAYDDQTSQ 120
 DB |||||
 62 IAKPPPPMGLLGPTIOAEVYDVTVTILKNMASHVSLHAGVSVYKASEGAYDDQTSQ 121
 QY 121 REKDDKYPFGSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 180
 DB |||||
 122 REKDDKYPFGSGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 181
 QY 181 EGS�AKEKTQTLHKFILLFAVDEGKSHWSETKSLMDQDRDAASARAPKMTVNGYVNR 240
 DB |||||
 182 EGS�AKEKTQTLHKFILLFAVDEGKSHWSETKSLMDQDRDAASARAPKMTVNGYVNR 241
 QY 241 SLPLGIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVNRHROASLEISPTTFLTAQTL 300
 DB |||||
 242 SLPLGIGCHRSVYWHVIGMGTTPVHSIFLEGHTFLVNRHROASLEISPTTFLTAQTL 301
 QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPPEQLRMKNNEEABDYDDDLTDSEMDVVRP 360
 DB |||||
 302 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEPPEQLRMKNNEEABDYDDDLTDSEMDVVRP 361
 QY 361 DDDNSPSFIQIRSVAKHPKTVVHYIAAEEDWDYAPLAVLAPDDRSYKSOYVINGPQIRG 420
 DB |||||
 362 DDDNSPSFIQIRSVAKHPKTVVHYIAAEEDWDYAPLAVLAPDDRSYKSOYVINGPQIRG 421
 QY 421 RYKVKRFMAYTDETFKTREAIQHESGILGPLLYGEGVDTLIIIFKNQASRPYNYPRGI 480
 DB |||||
 422 RYKVKRFMAYTDETFKTREAIQHESGILGPLLYGEGVDTLIIIFKNQASRPYNYPRGI 481
 QY 481 TDVRPLYSRRLPKGVKHLKDPPIIPGIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 540
 DB |||||
 482 TDVRPLYSRRLPKGVKHLKDPPIIPGIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNM 541
 QY 541 RDLASGLIGPLLCYKESVDQGNQINSDKRNVLFSVFDENRSWYLTENIQRFLNPAG 600
 DB |||||
 542 RDLASGLIGPLLCYKESVDQGNQINSDKRNVLFSVFDENRSWYLTENIQRFLNPAG 601
 QY 601 VQLEDPEFQASNIMHSINGVYFDSIQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 660
 DB |||||
 602 VQLEDPEFQASNIMHSINGVYFDSIQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 661
 QY 661 KWTYEDTLTLPFPGSETVFMSENPGMLWGLCHNSDFRNRGWTALLKVSSCDKNTGYDE 720
 DB |||||

Db 662 KAVYEDTLTLPFSGETVMSMENPGLWILGCHNSDFRNRMGTALLKVSSCDKXTGDIYE 721
QY 721 DSYEDISAYLLSKNNAIEPRFSQNPPLVKRHRERITRTTLODQEEIDYDDTISVEMKK 780
Db 722 DSYEDISAYLLSKNNAIEP-----REITRTTLODQEEIDYDDTISVEMKK 767
QY 781 EDFDIYDEENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVPOFKKV 840
Db 768 EDFDIYDEENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVPOFKKV 827
QY 841 FOEFTDGSFTQPLYRGELNEHLGLLPYIRAEVDENIMVTFRNQASRPYSFYSSLSIYEE 900
Db 828 FOEFTDGSFTQPLYRGELNEHLGLLPYIRAEVDENIMVTFRNQASRPYSFYSSLSIYEE 887
QY 901 DORQGAEPKRNFKVNETKTYFWKVQHMAPYKOBDFCKAWAYFSDVDLEKDVHSGLIGP 960
Db 888 DORQGAEPKRNFKVNETKTYFWKVQHMAPYKOBDFCKAWAYFSDVDLEKDVHSGLIGP 947
QY 961 LLVCHTNTLNPAHGRQVTVQEPALPFTIFDETKSMYFTENMERNCRAPCNIOMEDPTKE 1020
Db 948 LLVCHTNTLNPAHGRQVTVQEPALPFTIFDETKSMYFTENMERNCRAPCNIOMEDPTKE 1007
QY 1021 NYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIHPSGHVFTVRKXEYKMA 1080
Db 1008 NYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIHPSGHVFTVRKXEYKMA 1067
QY 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNKCOTPLGMASGHIRD 1140
Db 1068 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNKCOTPLGMASGHIRD 1127
QY 1141 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVYDILLAPMIHGIKTQGAROKFS 1200
Db 1128 FOITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVYDILLAPMIHGIKTQGAROKFS 1187
QY 1201 SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFGNVDSGIXHNIENPPIIARYIRLHPT 1260
Db 1188 SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFGNVDSGIXHNIENPPIIARYIRLHPT 1247
QY 1261 HYSIRSTLRMELMGCDLNSCMLPGWESKAI SDAQITASSYFTNMFTWSPSKARHLQG 1320
Db 1248 HYSIRSTLRMELMGCDLNSCMLPGWESKAI SDAQITASSYFTNMFTWSPSKARHLQG 1307
QY 1321 RSNARWPQVNNPKWLQVDFQKTMKVTVTTQGVKSLLTSMYVKEFLISSODGHQWTLF 1380
Db 1308 RSNARWPQVNNPKWLQVDFQKTMKVTVTTQGVKSLLTSMYVKEFLISSODGHQWTLF 1367
QY 1381 FQNGKVPQGNQDSFTPVNSLDPLLRYLRIHPQSWVHOIALRMEVLGCEAODLY 1438
Db 1368 FQNGKVPQGNQDSFTPVNSLDPLLRYLRIHPQSWVHOIALRMEVLGCEAODLY 1425

Search completed: December 9, 2003, 17:02:12
Job time : 68 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 17:04:05 ; Search time 22 Seconds
(without alignments)
2765.592 Million cell updates/sec

Title: US-10-047-257-1
Perfect score: 7691
Sequence: 1 ATRRYLGAVELSNDYMSQD.....WVHQIALRMEVLGCEAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/aaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/aaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/aaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/aaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/aaa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7691	100.0	1438	4	US-09-209-916-1
2	7674	99.8	1471	1	US-08-683-8398-3
3	7434	96.7	1661	2	US-08-882-083-2
4	7434	96.7	1661	3	US-08-558-107-2
5	7434	96.7	1661	3	US-09-243-539-2
6	7234	94.1	2332	1	US-07-864-004B-4
7	7234	94.1	2332	1	US-08-251-937A-4
8	7234	94.1	2332	1	US-08-212-133A-2
9	7234	94.1	2332	1	US-08-474-503-2
10	7234	94.1	2332	2	US-08-670-707A-2
11	7234	94.1	2332	3	US-09-037-601-2
12	7234	94.1	2332	3	US-09-315-179-2
13	7234	94.1	2332	4	US-09-523-656-2
14	7234	94.1	2332	5	PCT-US93-03275-4
15	7234	94.1	2332	5	PCT-US94-13200-2
16	7234	94.1	2351	1	US-08-121-202-2
17	7227	94.0	2351	1	US-08-366-851A-2
18	7225	93.9	2351	6	5422260-1
19	7224	93.9	2332	1	US-08-276-594A-2
20	7221	93.9	2351	6	5171844-2
21	7195	93.6	2332	3	US-09-324-867-3
22	6560	85.3	1467	4	US-09-523-656-38
23	6489	84.4	1443	3	US-08-670-707A-39
24	6489	84.4	1443	3	US-09-037-601-39
25	6489	84.4	1443	3	US-09-315-179-39
26	6272	81.5	2343	3	US-09-324-867-2
27	6227	81.0	2133	2	US-08-670-707A-37

28	6227	81.0	2133	3	US-09-037-601-37	Sequence 37, Appl
29	6227	81.0	2133	4	US-09-315-179-37	Sequence 37, Appl
30	6227	81.0	2133	4	US-09-523-656-30	Sequence 30, Appl
31	6192	80.5	2319	1	US-08-212-133A-8	Sequence 8, Appl
32	6192	80.5	2319	2	US-08-474-503-6	Sequence 6, Appl
33	6192	80.5	2319	2	US-08-670-707A-6	Sequence 6, Appl
34	6192	80.5	2319	3	US-09-037-601-6	Sequence 6, Appl
35	6192	80.5	2319	4	US-09-315-179-6	Sequence 6, Appl
36	6192	80.5	2319	4	US-09-523-656-28	Sequence 28, Appl
37	6192	80.5	2319	5	PCT-US94-13200-6	Sequence 6, Appl
38	6187.5	80.5	2115	3	US-09-324-867-5	Sequence 5, Appl
39	6147	79.9	2304	3	US-09-324-867-4	Sequence 4, Appl
40	2444.5	31.8	541	1	US-08-121-202-4	Sequence 4, Appl
41	2369	30.8	2183	3	US-08-746-111-5	Sequence 5, Appl
42	1726	22.4	368	1	US-08-212-133A-6	Sequence 6, Appl
43	1726	22.4	368	1	US-08-474-503-4	Sequence 4, Appl
44	1726	22.4	368	2	US-08-670-707A-4	Sequence 4, Appl
45	1726	22.4	368	3	US-09-037-601-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-209-916-1
; Sequence 1, Application US/09209916
; Patent No. 6358703
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/09/209,916
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-09-209-916-1

Query Match	100.0%;	Score 7691;	DB 4;	Length 1438;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1438;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATRRYLGAVELSNDYMSQD/GELPVDARFPVRPKSPFNTSVYKTLFVEFVHLFN	60	
Db	1	ATRRYLGAVELSNDYMSQD/GELPVDARFPVRPKSPFNTSVYKTLFVEFVHLFN	60	
QY	61	IAKPRPPMGLGPTIQAEVYDVTVTIKNMAHPVSLHVGVSYWKASGAEVDDQTSQ	120	
Db	61	IAKPRPPMGLGPTIQAEVYDVTVTIKNMAHPVSLHVGVSYWKASGAEVDDQTSQ	120	
QY	121	REKDDKVPFGSHTYVQVLEKNGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR	180	
Db	121	REKDDKVPFGSHTYVQVLEKNGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR	180	
QY	181	EGSLAKETQPLHFKIFLLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTYGVNVR	240	
Db	181	EGSLAKETQPLHFKIFLLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTYGVNVR	240	
QY	241	SLPGLIGCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPTFTLTALL	300	
Db	241	SLPGLIGCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPTFTLTALL	300	
QY	301	MDLQQLFCHSHSHQHDGMEAVYKVDSCPEEPOLRMNNNEAEYDDDLTDSMDVYRF	360	
Db	301	MDLQQLFCHSHSHQHDGMEAVYKVDSCPEEPOLRMNNNEAEYDDDLTDSMDVYRF	360	

Db 301 MDLGGQLLCHTSSHQHGDGMEAYVVKVDSCEPQPMKQNEAEADYDDDLTDSMDVVR 360
QY 361 DDNSPSFIQIRSAVKHPKTHVYIAAEEEDWDYAPLVLPDDRKYKQYLNNQFQRIG 420
Db 361 DDNSPSFIQIRSAVKHPKTHVYIAAEEEDWDYAPLVLPDDRKYKQYLNNQFQRIG 420
QY 421 RKYKKVRWAYTDETFKTRATQHESGILGPLLYGEGVDTLLIIFKQASRPYNIYPHGI 480
Db 421 RKYKKVRWAYTDETFKTRATQHESGILGPLLYGEGVDTLLIIFKQASRPYNIYPHGI 480
QY 481 TDVRLYSRRLPGVXHLKDFPILPGEIFKIKWTVVDEGPTKSDPRCLTRYSSVWME 540
Db 481 TDVRLYSRRLPGVXHLKDFPILPGEIFKIKWTVVDEGPTKSDPRCLTRYSSVWME 540
QY 541 RLASGLIGPLLCYKESVDQRQNMDSKRNVLFSVFDENRNSWYLTENIQRLPNPAG 600
Db 541 RLASGLIGPLLCYKESVDQRQNMDSKRNVLFSVFDENRNSWYLTENIQRLPNPAG 600
QY 601 VOLEDEFCASNMHSINGVVDLSQLSVCLHEVAYWYILSGAOTDFLSVPSGYTFKH 660
Db 601 VOLEDEFCASNMHSINGVVDLSQLSVCLHEVAYWYILSGAOTDFLSVPSGYTFKH 660
QY 661 KMVYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSCKDKNGDYIE 720
Db 661 KMVYEDTLTLPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSCKDKNGDYIE 720
QY 721 DSYEDISAYLLSKNNAIEPRSFQNPVLPKQHOREITRTLQSDQBEIDYDDTISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIEPRSFQNPVLPKQHOREITRTLQSDQBEIDYDDTISVEMKK 780
QY 781 EDFDIYDEENQSPRQFQKTRHYFTAAVERLMDYGMSSSPHVRRAQSGVPOPKVV 840
Db 781 EDFDIYDEENQSPRQFQKTRHYFTAAVERLMDYGMSSSPHVRRAQSGVPOPKVV 840
QY 841 FQFTDGSFTQPLRYGELNEHLGLGPIYRAEVEDNMVTFRNQASRPYSFSSLSIYEE 900
Db 841 FQFTDGSFTQPLRYGELNEHLGLGPIYRAEVEDNMVTFRNQASRPYSFSSLSIYEE 900
QY 901 DQOQGAEPKPNFKNETKTYFKVQHNAPTKDEFCCKAWAYFSDVDLEKDVHSLGIP 960
Db 901 DQOQGAEPKPNFKNETKTYFKVQHNAPTKDEFCCKAWAYFSDVDLEKDVHSLGIP 960
QY 961 LLVCHTNTLNPAGROVTVQEFALFTTFDETSKSWTFENMERNCRAPCNQIOMEDTPE 1020
Db 961 LLVCHTNTLNPAGROVTVQEFALFTTFDETSKSWTFENMERNCRAPCNQIOMEDTPE 1020
QY 1021 NYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEYYMA 1080
Db 1021 NYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSGNSNENIHSIFSGHVTVRKKEYYMA 1080
QY 1081 LYNLYPGVETVEMLPKAGIWEVCLIGELHAGHASTLFLVYSNKCOTPLGWSGHIRD 1140
Db 1081 LYNLYPGVETVEMLPKAGIWEVCLIGELHAGHASTLFLVYSNKCOTPLGWSGHIRD 1140
QY 1141 FQITASQXGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFS 1200
Db 1141 FQITASQXGQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFS 1200
QY 1201 SLVYSOFIIMSLDGKQWOTYRGNSTGLMVFPGNDSSGIXKINIPNPIIARYIRLHPT 1260
Db 1201 SLVYSOFIIMSLDGKQWOTYRGNSTGLMVFPGNDSSGIXKINIPNPIIARYIRLHPT 1260
QY 1261 HYSIRSTRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFATWSPKARLHLOQ 1320
Db 1261 HYSIRSTRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFATWSPKARLHLOQ 1320
QY 1321 RSNAPQVNPKNEMLOVDFQKTMKVTVTQGVKSLLTSMYKFEFLISSQDGHQWTLF 1380
Db 1321 RSNAPQVNPKNEMLOVDFQKTMKVTVTQGVKSLLTSMYKFEFLISSQDGHQWTLF 1380
QY 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLTLYLRIHPQSWHQIALRMEVLGCEAQDLY 1438
Db 1381 FQNGKVKVQGNQDSFTPVVNSLDPLLTLYLRIHPQSWHQIALRMEVLGCEAQDLY 1438

RESULT 2

US-08-683-839B-3
; Sequence 3, Application US/08683839B
; Patent No. 5744326
; GENERAL INFORMATION:
; APPLICANT: ILL, Charles . R. et al.
; TITLE OF INVENTION: Use of Viral Cis-Acting Post-Transcriptional
; TITLE OF INVENTION: Regulatory Sequences To Increase Expression of
; TITLE OF INVENTION: Introns Genes Containing Near-Consensus Splice Sites
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,839B
; FILING DATE: 11-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Remillard, Jane E.
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: TTI-138
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLESCALE TYPE: protein
US-08-683-839B-3
Query Match 99.8%; Score 7674; DB 1; Length 1471;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;
QY 1 ATRRYTLGAVELSWNDYMSDGLGELPVDARFPFRVPKSPFFNTSVYTKTLFVEFTVHLFN 60
Db 20 ATRRYTLGAVELSWNDYMSDGLGELPVDARFPFRVPKSPFFNTSVYTKTLFVEFTVHLFN 79
QY 61 IAKRPPPMWGLGPTIQAEVYDVTVTILKNMASHVPSLHVGVSVMKASEGAEYDDQTSQ 120
Db 80 IAKRPPPMWGLGPTIQAEVYDVTVTILKNMASHVPSLHVGVSVMKASEGAEYDDQTSQ 139
QY 121 REKEDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDYKOLNSGLIGALLVCR 180
Db 140 REKEDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLSHVLDYKOLNSGLIGALLVCR 199
QY 181 EGSIAKETQTLHKFILLPAVFDGKSWHSTKSLMDQDRDAASARAPKQHTVNGYNR 240
Db 200 EGSIAKETQTLHKFILLPAVFDGKSWHSTKSLMDQDRDAASARAPKQHTVNGYNR 259
QY 241 SLPLGLIGCHRSVYWHVIGMTTPVHSIFLEGHTFLVNRHQASLEISPTITLTAQTLL 300
Db 260 SLPLGLIGCHRSVYWHVIGMTTPVHSIFLEGHTFLVNRHQASLEISPTITLTAQTLL 319
QY 301 MDLGGQLLCHTSSHQHGDGMEAYVVKVDSCEPQPMKQNEAEADYDDDLTDSMDVVR 360
Db 320 MDLGGQLLCHTSSHQHGDGMEAYVVKVDSCEPQPMKQNEAEADYDDDLTDSMDVVR 379

QY 361 DDNSPSFQIRSVAKKPKTWHTYIAABEEDNDYAPLVLPDDRYSKYQYLANGPQRTG 420
Db 380 DDNSPSFQIRSVAKKPKTWHTYIAABEEDNDYAPLVLPDDRYSKYQYLANGPQRTG 439
QY 421 RYKVKRFPAYTDTETFKTRAIQHESGILGPLYLGEVGTLLIIFKNQASRPYNIYPHGI 480
Db 440 RYKVKRFPAYTDTETFKTRAIQHESGILGPLYLGEVGTLLIIFKNQASRPYNIYPHGI 499
QY 481 TDVRLPLSRRLPKGVKHLKDFPILPGEIIFKYKVTVTVEGPTKSDPRCLTRYISFVNME 540
Db 500 TDVRLPLSRRLPKGVKHLKDFPILPGEIIFKYKVTVTVEGPTKSDPRCLTRYISFVNME 559
QY 541 RDLASGLIGLLLCYKESVDORGNQIMSDKRNILFSVFDENRSLWTENIQRLPNPAG 600
Db 560 RDLASGLIGLLLCYKESVDORGNQIMSDKRNILFSVFDENRSLWTENIQRLPNPAG 619
QY 601 VQLEDPEFQASNIHMSINGVYFDSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
Db 620 VQLEDPEFQASNIHMSINGVYFDSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 679
QY 661 KMVYEDTLTLPFSGETVMSMENRGLWILGCHNSDFRNGMTALLKVSCKNTGDYIE 720
Db 680 KMVYEDTLTLPFSGETVMSMENRGLWILGCHNSDFRNGMTALLKVSCKNTGDYIE 739
QY 721 DSYEDISAYLLSKNNAIEPRFSQN-----PPVLKRRHOREITRTTLOSQOE 766
Db 740 DSYEDISAYLLSKNNAIEPRFSQNSHPSTROKQFNATPPVLKRRHOREITRTTLOSQOE 799
QY 767 EIDYDDTISVEMKEDFDIYDEBENSQPSFQKTRHYTIAAVERLWDYCMSSSPHVLN 826
Db 800 EIDYDDTISVEMKEDFDIYDEBENSQPSFQKTRHYTIAAVERLWDYCMSSSPHVLN 859
QY 827 RAQSGSVOPKVKVVOEFTDGSFTQPLRGELNEHLGILGPYIRAVEDNIMVTFRNQAS 886
Db 860 RAQSGSVOPKVKVVOEFTDGSFTQPLRGELNEHLGILGPYIRAVEDNIMVTFRNQAS 919
QY 887 RPYSFYSSLYSIEBDOQAEPRKNFKNETKYFWKQHMATPKDFDCKAWAYFD 946
Db 920 RPYSFYSSLYSIEBDOQAEPRKNFKNETKYFWKQHMATPKDFDCKAWAYFD 979
QY 947 VLEKDVHSLGILPLVCHTNTLNPAHQSVTVQEFALFTTFIDETKSWYFTENMERNC 1006
Db 980 VLEKDVHSLGILPLVCHTNTLNPAHQSVTVQEFALFTTFIDETKSWYFTENMERNC 1039
QY 1007 APCNIQWEDPTKENVRFHAIINGYIMDTLPLGLVMAODQIRWYLLSMGNSNENIHSFSG 1066
Db 1040 APCNIQWEDPTKENVRFHAIINGYIMDTLPLGLVMAODQIRWYLLSMGNSNENIHSFSG 1099
QY 1067 HVFTVRKKEEYKQALYNLYPGVPETVEMLPKSGAGIWRVECLIGELHAGNSTLFLVYSNK 1126
Db 1100 HVFTVRKKEEYKQALYNLYPGVPETVEMLPKSGAGIWRVECLIGELHAGNSTLFLVYSNK 1159
QY 1127 COTPLGMSGHIRDFOITASQYQWAPKLARLHSGSINASTKEPPEWIKVDLLAPMI 1186
Db 1160 COTPLGMSGHIRDFOITASQYQWAPKLARLHSGSINASTKEPPEWIKVDLLAPMI 1219
QY 1187 IHGKITQGARQFSSLYISQFIIMYSLDGKKQTVRGNSTGTLMVFFGNVDSGKHNIF 1246
Db 1220 IHGKITQGARQFSSLYISQFIIMYSLDGKKQTVRGNSTGTLMVFFGNVDSGKHNIF 1279
QY 1247 NPPIIARYRLHPTHTYSIRSTLRMELMGCDLNSCSMPLGWSKASDAQITASSYFTNMF 1306
Db 1280 NPPIIARYRLHPTHTYSIRSTLRMELMGCDLNSCSMPLGWSKASDAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHSGSNAREPOVNPKEWLQYDFQKTMKVTGVTTCQVKSLLTSMYVKEP 1366
Db 1340 ATWSPSKARLHSGSNAREPOVNPKEWLQYDFQKTMKVTGVTTCQVKSLLTSMYVKEP 1399
QY 1367 LISSSQDGHQWTLFFQNGKVKVYFQGNQDSFTPVVNSLDPPLATRYLRHPQSWHQAIALR 1426
Db 1400 LISSSQDGHQWTLFFQNGKVKVYFQGNQDSFTPVVNSLDPPLATRYLRHPQSWHQAIALR 1459

QY 1427 MEVLGCEADLY 1438
Db 1460 MEVLGCEADLY 1471
RESULT 3
US-08-882-083-2
; Sequence 2, Application US/08882083
; Patent No. 5869292
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,083
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/558,107
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-882-083-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;
QY 1 ATRRYYLGAVELSNDYMQSDLGELPVDARFPFPPVPKSPFFNTSVVYKKTFLVFETVHLFN 60
Db 20 ATRRYYLGAVELSNDYMQSDLGELPVDARFPFPPVPKSPFFNTSVVYKKTFLVFETVHLFN 79
QY 61 IAKRPPPMGLIGTTIQAEVYDVTWITLKNASHPVSLHVGVSWKASGAEYDDQTSQ 120
Db 80 IAKRPPPMGLIGTTIQAEVYDVTWITLKNASHPVSLHVGVSWKASGAEYDDQTSQ 139
QY 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 140 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGSIAKKTQTLHKPILLFAVFDGKSWHSSTKNSLMQDRDAASARAMPKQHTVGVYNR 240
Db 200 EGSIAKKTQTLHKPILLFAVFDGKSWHSSTKNSLMQDRDAASARAMPKQHTVGVYNR 259
QY 241 SLPLGICHKRSVYWHVIGMGTTPVEHSIFLEGHTFLVRNHRQASLSIPTFTLTAQTL 300
Db 260 SLPLGICHKRSVYWHVIGMGTTPVEHSIFLEGHTFLVRNHRQASLSIPTFTLTAQTL 319
QY 301 MDLQQLFLFCHISSHQHGMGEAYVKVDSCEPPEQLRMKNNEEADYDDDLTDSMDVVRF 360

Db 320 MDLQGLLFCHISSHGDGMEAYVKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVRP 379
QY 361 DDNSPSFIQIRSAVKHPKWTWHYIAAEEDWDYAPLVAPDDRYSKQYLNNGPORIG 420
Db 380 DDNSPSFIQIRSAVKHPKWTWHYIAAEEDWDYAPLVAPDDRYSKQYLNNGPORIG 439
QY 421 RYKVKRPMAYTDETPKTRAIQHESGILGPLLAYGEVGTLLIFRQASRPYNIYPHGI 480
Db 440 RYKVKRPMAYTDETPKTRAIQHESGILGPLLAYGEVGTLLIFRQASRPYNIYPHGI 499
QY 481 TDVRPLYSRLPKGVHLKDPPLRGEIFKYKNTVTVEDGPTKSDPRCLTRYSSFWNME 540
Db 500 TDVRPLYSRLPKGVHLKDPPLRGEIFKYKNTVTVEDGPTKSDPRCLTRYSSFWNME 559
QY 541 RDLASGLIGPLLCYKESVDQRNQIMSKRNVLFSVFDENRSLWTENIQRFENPAG 600
Db 560 RDLASGLIGPLLCYKESVDQRNQIMSKRNVLFSVFDENRSLWTENIQRFENPAG 619
QY 601 VQLEDPEFOASNMHSGINGVFDLSQLSVCLHVAWYVILSIGAQTDFLSVFFSGYTFKH 660
Db 620 VQLEDPEFOASNMHSGINGVFDLSQLSVCLHVAWYVILSIGAQTDFLSVFFSGYTFKH 679
QY 661 KMVYEDTLTLPFSGETVFMENPGIWLIGCHNSDFRNGMTALLKVVSC---DKNTGD 717
Db 680 KMVYEDTLTLPFSGETVFMENPGIWLIGCHNSDFRNGMTALLKVVSCPEGEEDD 739
QY 718 YY-----EDSYEDISAYLLSKNNAIBPRSF----- 742
Db 740 YLDLEKIFSEDDYIDI-----VDSLIERSFSQNSRHPSTKQKFNATTIPENDIBKT 793
QY 743 ----- 742
Db 794 DPMFAHRTMPKTIQNVSSDMLMLLRQSPTHGLSLDLQEAKEYTFSDPSGADSN 853
QY 743 ----- 742
Db 854 SLSEWTHFRPQLHSGDMVTFPSGLQLRLNEKLGTADPLAWNHYGQIPKEEKWSE 913
QY 743 -----SQNPPVLK 750
Db 914 KSPEKTAFFKKTILSINACESHATAINEGONKEPEIEVTWAKQGERLCSQNPVLK 973
QY 751 RHOREITRTTQSDQBEIDYDDTISVEMKEDFDIYDEENQSPRSFQKKTHTYTAAYE 810
Db 974 RHOREITRTTQSDQBEIDYDDTISVEMKEDFDIYDEENQSPRSFQKKTHTYTAAYE 1033
QY 811 RLWDYGMSSSPHVLNRNAQSGSVQPKVVFQEFDTGSGFTQPLRGELNEHLGLLCPYIR 870
Db 1034 RLWDYGMSSSPHVLNRNAQSGSVQPKVVFQEFDTGSGFTQPLRGELNEHLGLLCPYIR 1093
QY 871 AEVEDNIMVTFNRQASRPYSFVSSLSISYEDQOQABPRKFNKNETTYFWKQVHMA 930
Db 1094 AEVEDNIMVTFNRQASRPYSFVSSLSISYEDQOQABPRKFNKNETTYFWKQVHMA 1153
QY 931 PKDEPDKAWAYFSDVLEKDVHSLIGLPLVCHTNTLNPAGHQVTVQEPALFTIFD 990
Db 1154 PKDEPDKAWAYFSDVLEKDVHSLIGLPLVCHTNTLNPAGHQVTVQEPALFTIFD 1213
QY 991 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTLPLGMAQDQIRWYL 1050
Db 1214 ETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTLPLGMAQDQIRWYL 1273
QY 1051 LSWGSENIHSTHESGHVTFVRKKEBKWALYNLPGVETVEMLPKAGIWRVCLIGE 1110
Db 1274 LSWGSENIHSTHESGHVTFVRKKEBKWALYNLPGVETVEMLPKAGIWRVCLIGE 1333
QY 1111 HLHAGSTLFLVYSNKCOTPLGNASGHIRDFQITASGOYGOWAPKLARLHYSGSNAST 1170
Db 1334 HLHAGSTLFLVYSNKCOTPLGNASGHIRDFQITASGOYGOWAPKLARLHYSGSNAST 1393
QY 1171 KEPPFWIKVDLLAPMIHGIKIQGARQKPSLYISQFIIMYSLDGKKQWYRGNSTGLM 1230

Db 1394 KEPPFWIKVDLLAPMIHGIKIQGARQKPSLYISQFIIMYSLDGKKQWYRGNSTGLM 1453
QY 1231 VFPGNVDSGIGHNIFNPPIIARYIRLHPHTHYSIBSTLMEMLMGCDLNSCSHPLGNESKA 1290
Db 1454 VFPGNVDSGIGHNIFNPPIIARYIRLHPHTHYSIBSTLMEMLMGCDLNSCSHPLGNESKA 1513
QY 1291 ISDAQITASSYFTNMFEATMSPSKARLHLQGRSNARMPQVNNPKWLQVDFOKTMKVTVGT 1350
Db 1514 ISDAQITASSYFTNMFEATMSPSKARLHLQGRSNARMPQVNNPKWLQVDFOKTMKVTVGT 1573
QY 1351 TQGVKSLTSMYKFLPFISSSQDGHQWTLFPQNGKVKVPOGQNDSTFTPVNSLDPPLTR 1410
Db 1574 TQGVKSLTSMYKFLPFISSSQDGHQWTLFPQNGKVKVPOGQNDSTFTPVNSLDPPLTR 1633
QY 1411 YLRIHPQSWHQAIALRMEVLGCEAODLY 1438
Db 1634 YLRIHPQSWHQAIALRMEVLGCEAODLY 1661

RESULT 4

US-08-558-107-2
; Sequence 2, Application US/08558107
; Patent No. 5910481
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/558.107
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-558-107-2

Query Match 96.7%; Score 7434; DB 2; Length 1661;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;
QY 1 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
Db 20 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 79
QY 61 IAKRPPWMLGLGPTIOAEVYDVTWVITLKNASHPVSILHVGVSYKASGAEYDDQTSQ 120
Db 80 IAKRPPWMLGLGPTIOAEVYDVTWVITLKNASHPVSILHVGVSYKASGAEYDDQTSQ 139
QY 121 REKEDDKVFPQSGSTTYWQVLEKNGPMASDPLCLTYSVLSHVDLAVKLSGLIGALLVCR 180

Db 140 REKEDKVPFGSGSHYVVOVLKENGPMASDPCLCLATYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGSIAKEKQTLLKFKILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMTWVNGYVNR 240
Db 200 EGSIAKEKQTLLKFKILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMTWVNGYVNR 259
QY 241 SLPLGICHRKSVYMHVIGMTTPEVHSIFLBGHTFLVRNHRQASLEISPIITFLTAQTLL 300
Db 260 SLPLGICHRKSVYMHVIGMTTPEVHSIFLBGHTFLVRNHRQASLEISPIITFLTAQTLL 319
QY 301 MDLQGFLLCHLSSHODGMEAVKVDSCPEPQRLMKNEAEYDDDLTSEMDVVRP 360
Db 320 MDLQGFLLCHLSSHODGMEAVKVDSCPEPQRLMKNEAEYDDDLTSEMDVVRP 379
QY 361 DDNSPSFIQIRSAVKKPKTWHYIAABEEDWDYAPLVAPLAPDRSYKQYLNNGPQIRG 420
Db 380 DDNSPSFIQIRSAVKKPKTWHYIAABEEDWDYAPLVAPLAPDRSYKQYLNNGPQIRG 439
QY 421 RYKVKVRMAYTDETFKTEAIOHESGILGPLLYGEVGTLLIIFKNQASRPYNTYPHGI 480
Db 440 RYKVKVRMAYTDETFKTEAIOHESGILGPLLYGEVGTLLIIFKNQASRPYNTYPHGI 499
QY 481 TVRPLYSRRLPKGVKHLDFPILPGEIPIKYKMTVTVEDGPKSDPRCLTRYYSFVNME 540
Db 500 TVRPLYSRRLPKGVKHLDFPILPGEIPIKYKMTVTVEDGPKSDPRCLTRYYSFVNME 559
QY 541 RDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNPNPAG 600
Db 560 RDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFNPNPAG 619
QY 601 VOLEDEPOASIMHSINGYVDSLSQSVCLHEVAYWYILSGAOTDFLSVFSGYTPKH 660
Db 620 VOLEDEPOASIMHSINGYVDSLSQSVCLHEVAYWYILSGAOTDFLSVFSGYTPKH 679
QY 661 KAVYEDTLTLPFSGETVFMSENFGILGCHNSDFRNGMTALLKVSSC---DKNCD 717
Db 680 KAVYEDTLTLPFSGETVFMSENFGILGCHNSDFRNGMTALLKVSSCIPGEBEDD 739
QY 718 YY-----EUSYEDISAYLSKXNAIEPRGF----- 742
Db 740 YLDLEKIFSEDDYIDI-----VDSLIEPRFSQNSRHPSTRQKQFNATTIPENDIEKT 793
QY 743 ----- 742
Db 794 DPWFARHTPMPKIQNVSSDMLMLLRQSPTRPHGLSLDLQEAKEYTFSDPSGAIDSN 853
QY 743 ----- 742
Db 854 SLSEWTHPRQLHSGDMVFTPESGIQLRLNEKLGTTADPLANDNHYGTQIPKEWKSQE 913
QY 743 -----SQNPPVLK 750
Db 914 KSPEKTAFFKKDVTILSLNACESHAAIAINEGQNKPEIEVTWAKQGRTERLCSQNPVLK 973
QY 751 RHOREITRTTLOSDEEIDYDITISVEMKEDFDLYDRDENOSPEFQKTRHYFAIVE 810
Db 974 RHOREITRTTLOSDEEIDYDITISVEMKEDFDLYDRDENOSPEFQKTRHYFAIVE 1033
QY 811 RLMDYGMSSPHVLNRNRAQSGSVPOFKVVOEFTDGSFTQPLYRGELNEHLGLGPYIR 870
Db 1034 RLMDYGMSSPHVLNRNRAQSGSVPOFKVVOEFTDGSFTQPLYRGELNEHLGLGPYIR 1093
QY 871 AEVEDNIWTFNRQASRPYSFSSLSIYSEDOQQAEPKQFVKVKNENKTYFWKQVHMA 930
Db 1094 AEVEDNIWTFNRQASRPYSFSSLSIYSEDOQQAEPKQFVKVKNENKTYFWKQVHMA 1153
QY 931 PTKDEFDCWAYFSDVLEKDVHSLGILCPILLVCHTNTLNPAKGSQVTVQSFALPFTIFD 990
Db 1154 PTKDEFDCWAYFSDVLEKDVHSLGILCPILLVCHTNTLNPAKGSQVTVQSFALPFTIFD 1213
QY 991 ETKSGYFTENMERNCRAPCNIQMEDPTPKENYRFAINGYIMDTPLGLVMAQDQIRWYL 1050
Db 1214 ETKSGYFTENMERNCRAPCNIQMEDPTPKENYRFAINGYIMDTPLGLVMAQDQIRWYL 1273

QY 1051 LSMGSENIHSHFSGHVFTVRKKEEYKMALNLYPGVFETVEMLPKAGIWRVCECLIGE 1110
Db 1274 LSMGSENIHSHFSGHVFTVRKKEEYKMALNLYPGVFETVEMLPKAGIWRVCECLIGE 1333
QY 1111 HLHAGMSTLFLVYSNKKQTPPLGMASGHIRDFOITASGOYQWAPKLARLHYSGSINAWST 1170
Db 1334 HLHAGMSTLFLVYSNKKQTPPLGMASGHIRDFOITASGOYQWAPKLARLHYSGSINAWST 1393
QY 1171 KEPPSWIKVDLLAPMIHGIKTQGARQKPFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTILM 1230
Db 1394 KEPPSWIKVDLLAPMIHGIKTQGARQKPFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTILM 1453
QY 1231 VFFGNVDSGGIKHNIFFNPPIIARYIRLTHPTHYISIRSTMELMGDLNCSMPLGWESKA 1290
Db 1454 VFFGNVDSGGIKHNIFFNPPIIARYIRLTHPTHYISIRSTMELMGDLNCSMPLGWESKA 1513
QY 1291 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNARWPQVNNPKWLQVDFQKTMKVTGVT 1350
Db 1514 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNARWPQVNNPKWLQVDFQKTMKVTGVT 1573
QY 1351 TQGVKSLLTSMYKVEFLISSQDGHQWTLFFQNGKVKYFQGNQDSFTFVNSLDPPLLTR 1410
Db 1574 TQGVKSLLTSMYKVEFLISSQDGHQWTLFFQNGKVKYFQGNQDSFTFVNSLDPPLLTR 1633
QY 1411 YLRHPQSVWHQIALRMEVLGCEAQDLY 1438
Db 1634 YLRHPQSVWHQIALRMEVLGCEAQDLY 1661

RESULT 5

US-09-243-539-2
; Sequence 2, Application US/09243539
; Patent No. 6130203
; GENERAL INFORMATION:
; APPLICANT: VOORBERG, Johannes J.
; TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/243,539
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/558,107
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 30472/212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904135
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1661 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-243-539-2

Query Match 96.7%; Score 7434; DB 3; Length 1661;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1418; Conservative 4; Mismatches 10; Indels 216; Gaps 4;

QY 1 ATRRYVLGAVELSDWYQSDGLPVDARPPRPVPSFFPNTSVYKTLFVEFTVHLEN 60
DB 20 ATRRYVLGAVELSDWYQSDGLPVDARPPRPVPSFFPNTSVYKTLFVEFTVHLEN 79
QY 61 IAKPRPPMGLGPTTQAEYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ 120
DB 80 IAKPRPPMGLGPTTQAEYDVTWITLKNMASHPVSLHAGVSYWKASGEGAYDDQTSQ 139
QY 121 REKEDDKVPGGSHYVWVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
DB 140 REKEDDKVPGGSHYVWVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 199
QY 181 EGS LAKEKTQTLHKFTLLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
DB 200 EGS LAKEKTQTLHKFTLLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259
QY 241 SLPLGLICHRKSVYVWVGTTPEVHSIFLEHTEFLVNRHROASLEISPIITFLTAQTLL 300
DB 260 SLPLGLICHRKSVYVWVGTTPEVHSIFLEHTEFLVNRHROASLEISPIITFLTAQTLL 319
QY 301 MDLGOELLFCHISSHQHDMGAEVYKVDSCPEEPOLRMKNNEAEYDDDLTDSMDVVRP 360
DB 320 MDLGOELLFCHISSHQHDMGAEVYKVDSCPEEPOLRMKNNEAEYDDDLTDSMDVVRP 379
QY 361 DDNSPSFTQIRSAVKHPTWYHIAAEEEDWDYAPLAPDDRYSKYQLNNGPQRIG 420
DB 380 DDNSPSFTQIRSAVKHPTWYHIAAEEEDWDYAPLAPDDRYSKYQLNNGPQRIG 439
QY 421 RYKVKVFWAYTDETEKTEAHOESGILGLLYGEVGTLLIIFKQASRPNTYPHGI 480
DB 440 RYKVKVFWAYTDETEKTEAHOESGILGLLYGEVGTLLIIFKQASRPNTYPHGI 499
QY 481 TDVRPLYRRLPKGVKHLDFPILPGEIIFYKWTVTVEDGPKSDPRCLTRYYSFVWME 540
DB 500 TDVRPLYRRLPKGVKHLDFPILPGEIIFYKWTVTVEDGPKSDPRCLTRYYSFVWME 559
QY 541 ROLASGLIGPLICYKESVDQRGNQIMSDKRNVLFSVFDENRWSYLTENIQRFLEPNAG 600
DB 560 ROLASGLIGPLICYKESVDQRGNQIMSDKRNVLFSVFDENRWSYLTENIQRFLEPNAG 619
QY 601 VOLEDEFOASNMHSHINGVVPFSLQSVCLHEVAYWVILSIGACTDRLSVFSPGVTFKH 660
DB 620 VOLEDEFOASNMHSHINGVVPFSLQSVCLHEVAYWVILSIGACTDRLSVFSPGVTFKH 679
QY 661 KMVYEDTLTLPFSGTVMENPGLWILGCHNSDFRNRGTMALLKVSSC---DKNTGD 717
DB 680 KMVYEDTLTLPFSGTVMENPGLWILGCHNSDFRNRGTMALLKVSSCIPGEEDDD 739
QY 718 YI-----EDSYEDISAYLSKNAIEPRSF----- 742
DB 740 YLDLEKIFSEDDYDI-----VDSLEIPRFSQNSRHPSTRQKQFNATTIPENDIEKT 793
QY 743 ----- 742
DB 794 DPWFAPRTPMPKIQNVSSSLLMLLRQSPTPHGLSLSDLOEAKYETFFSDPSGAIDSN 853
QY 743 ----- 742
DB 854 SLSEMTHTPQLHSDMVFTPESGIQLRLNEKLGTTADPLAWNHNHYQTQIPKEBWSQE 913
QY 743 -----SONPPVLK 750
DB 914 KSPKTAFKKXDTIISLNACENHAIANEQONKPEIEVTWAKQGRTERLCSQPPVLK 973
QY 751 RHOREITRTTLOSDEIDYDDTISVEMCKEDFDIYDEDENSPQSFQKTRHYFAAVE 810
DB 974 RHOREITRTTLOSDEIDYDDTISVEMCKEDFDIYDEDENSPQSFQKTRHYFAAVE 1033
QY 811 RLWDYGMSSSPVLRNRAQSGSVPOFKKVVQFETDGSFTQPLRGELNEHLGLLGPYIR 870

DB 1034 RLWDYGMSSSPVLRNRAQSGSVPOFKKVVQFETDGSFTQPLRGELNEHLGLLGPYIR 1093
QY 871 ABEVDNIMVTNQAASRPYSFYSSLSIYSEEDORQCAEPRKNEFKVNETKTYFWKVOHMA 930
DB 1034 ABEVDNIMVTNQAASRPYSFYSSLSIYSEEDORQCAEPRKNEFKVNETKTYFWKVOHMA 1153
QY 931 PTKDOPDCAMAYFSDVDLEKDVHSLGLPLLVCHTNTLNPAHGROVTVQVEFALFFTFD 990
DB 1154 PTKDOPDCAMAYFSDVDLEKDVHSLGLPLLVCHTNTLNPAHGROVTVQVEFALFFTFD 1213
QY 991 ETKSWYFTENMERNCRAPCNQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQORIRYL 1050
DB 1214 ETKSWYFTENMERNCRAPCNQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQORIRYL 1273
QY 1051 LSMGSENIHSHFSGHVTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVVECLICE 1110
DB 1274 LSMGSENIHSHFSGHVTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVVECLICE 1333
QY 1111 HLHAGMSTLFLVYSNKKCTPLGMASGHIRDFOITASGOYQOWAPKLARLHYSGSINAWST 1170
DB 1334 HLHAGMSTLFLVYSNKKCTPLGMASGHIRDFOITASGOYQOWAPKLARLHYSGSINAWST 1393
QY 1171 KEPPSWIKVDLAPMIHGIKTQGAROKFESSIYISOFTIMYSLDGKKWOTYRGNSTGTLM 1230
DB 1394 KEPPSWIKVDLAPMIHGIKTQGAROKFESSIYISOFTIMYSLDGKKWOTYRGNSTGTLM 1453
QY 1231 VFFGNVDSSGINKHINFNPPIIARYIRLHPTHYSTRILRMELMGCDLNSCMPLGMSKA 1290
DB 1454 VFFGNVDSSGINKHINFNPPIIARYIRLHPTHYSTRILRMELMGCDLNSCMPLGMSKA 1513
QY 1291 ISDAQITASSYFTNMFATWSPSKARLHQGSNARWPOVNNPKWQLVDFOKTMKVGT 1350
DB 1514 ISDAQITASSYFTNMFATWSPSKARLHQGSNARWPOVNNPKWQLVDFOKTMKVGT 1573
QY 1351 TQGVKSLTSMYVKEFLSSQDGHQWTLFFQNGKVKVFOGNOQSFTPVNSLDPPLTR 1410
DB 1574 TQGVKSLTSMYVKEFLSSQDGHQWTLFFQNGKVKVFOGNOQSFTPVNSLDPPLTR 1633
QY 1411 YLRHPQSVHQAIALMEVLGCEAQDLY 1438
DB 1634 YLRHPQSVHQAIALMEVLGCEAQDLY 1661

RESULT 6
US-07-864-004B-4
; Sequence 4, Application US/07864004B
; Patent No. 5364771
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/07/864,004B
; FILING DATE: 07 APRIL 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508
 TELEFAX: 404-815-6555
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2332 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:
 ORGANISM: Homo sapien
 TISSUE TYPE: Liver cdna sequence
 US-07-864-004B-4

Query Match 94.1%; Score 7234; DB 1; Length 2332;
 Best local Similarity 61.7%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRRYLGVAVELSDWYQSDGLGELPVDARPPRPVPSFPPNTSVVYKTLFVEFTVHLFN 60
 DB |||||
 QY 1 ATRRYLGVAVELSDWYQSDGLGELPVDARPPRPVPSFPPNTSVVYKTLFVEFTVHLFN 60
 DB |||||
 QY 61 IAKPRPPMGLLGPTTQAEVYDVTVTITLKNMASHPVSLHAGVSYWKASEGAEYDDDTQ 120
 DB |||||
 QY 61 IAKPRPPMGLLGPTTQAEVYDVTVTITLKNMASHPVSLHAGVSYWKASEGAEYDDDTQ 120
 DB |||||
 QY 121 REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
 DB |||||
 QY 121 REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
 DB |||||
 QY 181 EGS LAKEKQTLLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPMHTVGVYNR 240
 DB |||||
 QY 181 EGS LAKEKQTLLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPMHTVGVYNR 240
 DB |||||
 QY 241 SLPLGLGCHRSYVHWVIGMTTPEVHSIFLEGTFLVRNHRQASLEISPTIFLTAQTLL 300
 DB |||||
 QY 241 SLPLGLGCHRSYVHWVIGMTTPEVHSIFLEGTFLVRNHRQASLEISPTIFLTAQTLL 300
 DB |||||
 QY 301 MDLQGFLLFCHSSHQHDGMEAYVVDSCPEEPQLMKNEEAEDYDDDLTDSEMDVVRP 360
 DB |||||
 QY 301 MDLQGFLLFCHSSHQHDGMEAYVVDSCPEEPQLMKNEEAEDYDDDLTDSEMDVVRP 360
 DB |||||
 QY 361 DDNSPSFTQIRSVAKKHPTWHTYIAAEEEDWDYAPLVLAEDDRSYKSYQLNNGPQRTG 420
 DB |||||
 QY 361 DDNSPSFTQIRSVAKKHPTWHTYIAAEEEDWDYAPLVLAEDDRSYKSYQLNNGPQRTG 420
 DB |||||
 QY 421 RYKVKRPMAYTDETPKTBIAQHEGILGPLLYGEGVDTLLIIFKNQASRPNTLYPHGI 480
 DB |||||
 QY 421 RYKVKRPMAYTDETPKTBIAQHEGILGPLLYGEGVDTLLIIFKNQASRPNTLYPHGI 480
 DB |||||
 QY 481 TDVRLYRRLPKGVKHLKDFPLPGELPKYKWTVTVEDGPTKSDPRCLTRYSSFVNME 540
 DB |||||
 QY 481 TDVRLYRRLPKGVKHLKDFPLPGELPKYKWTVTVEDGPTKSDPRCLTRYSSFVNME 540
 DB |||||
 QY 541 RLASGLIGPLATCYKESVDQRGNQIMSDKNVILFSPVEDNSRWYLTENIORFLNPAG 600
 DB |||||
 QY 541 RLASGLIGPLATCYKESVDQRGNQIMSDKNVILFSPVEDNSRWYLTENIORFLNPAG 600
 DB |||||
 QY 601 VQLEDPEFQASNMHMSINGVVFDSQLSVCLHEVAYWYILSIAQTDLSVFFSGVTFPH 660
 DB |||||
 QY 601 VQLEDPEFQASNMHMSINGVVFDSQLSVCLHEVAYWYILSIAQTDLSVFFSGVTFPH 660
 DB |||||
 QY 661 KMVEDTLTLFPFGSGTVMFMENPGLWILGCHNSDFRNRGMTALLKVVSSCDKNTGDIYE 720
 DB |||||
 QY 661 KMVEDTLTLFPFGSGTVMFMENPGLWILGCHNSDFRNRGMTALLKVVSSCDKNTGDIYE 720
 DB |||||
 QY 721 DSVEDISAVLLSKNNAIERPSF----- 742
 DB |||||
 QY 721 DSVEDISAVLLSKNNAIERPSFQNSRHPSTRKQKNATTIPENDIEKTDPMFAHRTMP 780
 DB |||||

QY 743 ----- 742
 DB 781 KIQWSSDLLMLLRQSPTPHGLSLDLQEAKEYETFDPPSPCAIDSNNSLSEMTFRPQ 840
 QY 743 ----- 742
 DB 841 LHSQDMVTPESGLQLRLNEKLGTTAATLKKLDFKVSSTNNLISTIPSDNLAAGTN 900
 QY 743 ----- 742
 DB 901 TSSLGPPMPVHYDSQDITTLFGKSSPLTESGGPILSEENNDKSLLESGLMNSQESSW 960
 QY 743 ----- 742
 DB 961 GKNVSTESGRLPKGRAGHPALLTKONALFKVSI SLLKNTKTSNNSATNRKTHIDGPSL 1020
 QY 743 ----- 742
 DB 1021 LIENSPSVQWNI LESTDEPKKVTPLIHDRMLDKNATLRLNHNENKTTSSKNMEMVQOK 1080
 QY 743 ----- 742
 DB 1081 KEGPIPPDAQNDMSFFKMLFLPESARWIORTHGKNLSLGGQSPKQLVSLGPKSVBG 1140
 QY 743 ----- 742
 DB 1141 QNFLSEKNKVVGKEFTKDVGLKEMVPFSSRNLF LTNLNLHNENTHNQEKKIQEIEK 1200
 QY 743 ----- 742
 DB 1201 KETLLQENVVLPIQHVTVTGRTKNFMKNVLLSTQWVEGSEYAGYAPVLQDFRSLNDSTNR 1260
 QY 743 ----- 742
 DB 1261 TKKHTAHSKKGEENLEGLGNQTKQIVEKYACTTRISPTSQNFVQTSRKBALKQPRL 1320
 QY 743 ----- 742
 DB 1321 PLEETELEBKRIIVDTSTQWKNMKHLPSTLTQIDYNEKKGAITQSPLSCLTRSHSI 1380
 QY 743 ----- 742
 DB 1381 PQANRSLPIAKVSSFPPIRPIYLTIRVL PQDNSSHLPAASVYKKDSGVQBSHFLOQAKK 1440
 QY 743 ----- 742
 DB 1441 NMLSLAILTLEMTGQREVGSLGTSATNSVTYKVENTVL PKPDLPKTSGKVELPKVHI 1500
 QY 743 ----- 742
 DB 1501 YOKDLPPETNSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
 QY 743 ----- 742
 DB 1561 LLDPLAWNHYGTQIPKESWKSQESPEKTAFFKDTTILSNACSNHAI AAINEGONKP 1620
 QY 743 ----- SONPVLKHHQREITRTTLOSQBEIDYDDTTISVEMKKEDFDIY 786
 DB 1621 EIEVTWAKQGRTERLCSQNPVVKRHHQREITRTTLOSQBEIDYDDTTISVEMKKEDFDIY 1680
 QY 787 DEDNQSPRSFQKTRHYFIAAVERLDYGNSSPHVLNPAQSGSVPOFKKVVFOEFTD 846
 DB 1681 DEDNQSPRSFQKTRHYFIAAVERLDYGNSSPHVLNPAQSGSVPOFKKVVFOEFTD 1740
 QY 847 GSFTQPLYRGLNHLGLGYPYIRAEVEDNIMVTFRQASRPYFYSSLI SYEEDQOQA 906
 DB |||||
 QY 1741 GSFTQPLYRGLNHLGLGYPYIRAEVEDNIMVTFRQASRPYFYSSLI SYEEDQOQA 1800
 DB |||||
 QY 907 EPRKNFVKPNETKTYFVKVQHMMAPTQDEDFCKAWAYPSVDLEKDVHSGLIGLPLVCHT 966
 DB |||||
 QY 1801 EPRKNFVKPNETKTYFVKVQHMMAPTQDEDFCKAWAYPSVDLEKDVHSGLIGLPLVCHT 1860
 DB |||||

QY 967 NTLNPAHQVTVQEPALFTTIFDETQSWYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1026
DB 1861 NTLNPAHQVTVQEPALFTTIFDETQSWYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1920
QY 1027 INGYIMDTPLGLVMAODQIRWYLLSWGSGNENIHSIFSGHYFTVTKKEEYKMALYNLYP 1086
DB 1921 INGYIMDTPLGLVMAODQIRWYLLSWGSGNENIHSIFSGHYFTVTKKEEYKMALYNLYP 1980
QY 1087 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMAASHIRDFOITAS 1146
DB 1981 GVPETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMAASHIRDFOITAS 2040
QY 1147 GYQGWAPKLARLHYSGSINASTKEPFSWKVDLLAPMIHGIKTQAGARQFSSLYISQ 1206
DB 2041 GYQGWAPKLARLHYSGSINASTKEPFSWKVDLLAPMIHGIKTQAGARQFSSLYISQ 2100
QY 1207 FIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGIRKHNIFNPPIIARYIRLHPTHSIRS 1266
DB 2101 FIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGIRKHNIFNPPIIARYIRLHPTHSIRS 2160
QY 1267 TLRMELMGCDLNSCMSPLGMESKASDAQITASSYFTNNMFATWSPSKARLHLQGRSNWR 1326
DB 2161 TLRMELMGCDLNSCMSPLGMESKASDAQITASSYFTNNMFATWSPSKARLHLQGRSNWR 2220
QY 1327 PQVNNPEKWLQVDQFQTKMTGVTGTVQVKSLLTSMYVKEFLISSQDGHQWTLFQNGKY 1386
DB 2221 PQVNNPEKWLQVDQFQTKMTGVTGTVQVKSLLTSMYVKEFLISSQDGHQWTLFQNGKY 2280
QY 1387 KVFQGNQDSFTPVNSLDPPLLTTRYLRIRHPQSVHQAIALRMEVLGCEAODLY 1438
DB 2281 KVFQGNQDSFTPVNSLDPPLLTTRYLRIRHPQSVHQAIALRMEVLGCEAODLY 2332

RESULT 7

US-08-251-937A-4
Sequence 4, Application US/08251937A
Patent No. 5583209
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/251,937A
FILING DATE: 31-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMU106DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6367
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cDNA sequence
US-08-251-937A-4
Query Match 94.1%; Score 7234; DB 1; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYYLCAVELSWDMQSDLGELPVDARFPVRPVPKSPFFNTSVYKKTFLVEFTVHLFN 60
DB 1 ATRRYYLCAVELSWDMQSDLGELPVDARFPVRPVPKSPFFNTSVYKKTFLVEFTVHLFN 60
QY 61 IAKPRPPMGLLGPPTQAEVYDVTVITLKNASHPVSLHAYGVSYWKASGAGYDDQTSQ 120
DB 61 IAKPRPPMGLLGPPTQAEVYDVTVITLKNASHPVSLHAYGVSYWKASGAGYDDQTSQ 120
QY 121 REKEDDKVPPGSGSHYVQVLKENGPMASDPLCLTYSVLSHVDLAKOLNSGLIGLAVCR 180
DB 121 REKEDDKVPPGSGSHYVQVLKENGPMASDPLCLTYSVLSHVDLAKOLNSGLIGLAVCR 180
QY 181 EGS�AKEKTQTLHKLFAVDFDEGKSWHSEKNSLMDRDAASARAWPKMHTVNGYVNR 240
DB 181 EGS�AKEKTQTLHKLFAVDFDEGKSWHSEKNSLMDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQLL 300
DB 241 SLPLGLIGCHRSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQLL 300
QY 301 MDLQGFLLFCHISSHOHGMEAYVKVDSCEPQOLRMKNNEAEYDDDLTDSMDVVRFF 360
DB 301 MDLQGFLLFCHISSHOHGMEAYVKVDSCEPQOLRMKNNEAEYDDDLTDSMDVVRFF 360
QY 361 DDNSPSFIQIRSVAKGHKPTWVHYIAAEEDMDYAPLVLAAPDDRYSQVLLNNGPQIRG 420
DB 361 DDNSPSFIQIRSVAKGHKPTWVHYIAAEEDMDYAPLVLAAPDDRYSQVLLNNGPQIRG 420
QY 421 RYKXKVRFMAYTDEFTKTRERAIQHESGILGELLVEGVDOTLLIFKQASBPNTYRPHGI 480
DB 421 RYKXKVRFMAYTDEFTKTRERAIQHESGILGELLVEGVDOTLLIFKQASBPNTYRPHGI 480
QY 481 TDVRLYSRRLPKGVKHLKDFPILPGEIFKTKWTVTVEDGPTKSDPRCLTRYYSFVNWE 540
DB 481 TDVRLYSRRLPKGVKHLKDFPILPGEIFKTKWTVTVEDGPTKSDPRCLTRYYSFVNWE 540
QY 541 RDLASGLIGPLLCYKESVDQGNQIMSDKKNVILFSVFDENRSHWLTENTORFLPNPAG 600
DB 541 RDLASGLIGPLLCYKESVDQGNQIMSDKKNVILFSVFDENRSHWLTENTORFLPNPAG 600
QY 601 VQLEDPERQASINMHSINGVYVDFSLQVCLLHEVAYVYLLSIGACTDFLSVFFSGSYTPKH 660
DB 601 VQLEDPERQASINMHSINGVYVDFSLQVCLLHEVAYVYLLSIGACTDFLSVFFSGSYTPKH 660
QY 661 KAVYEDTTLTLPFSGETVFMGMENPGLMWILCHNSDFNRGMGTALLKVSSCDKNTGYDE 720
DB 661 KAVYEDTTLTLPFSGETVFMGMENPGLMWILCHNSDFNRGMGTALLKVSSCDKNTGYDE 720
QY 721 DSYEDISAYLLSKXNAIEPRSP----- 742
DB 721 DSYEDISAYLLSKXNAIEPRSP----- 742
QY 743 ----- 742
DB 743 ----- 742
QY 781 KIQNVSSSLLMLLRQSFTHGLSLSLDLQEAKEYTFSDDFSGAIDSNNSLSEWTHFRPQ 840
DB 781 KIQNVSSSLLMLLRQSFTHGLSLSLDLQEAKEYTFSDDFSGAIDSNNSLSEWTHFRPQ 840
QY 743 ----- 742
DB 841 LHSQDMVFTPESGLQLRLNEKLGTTAATLKLKDFKVSSTSNLSTIPSDNLAAGTDN 900

```
QY 743 ----- 742
Db 901 TSSGLPSPVHVDSQDITLFGKSSPLTESGGPLSLEENNDKLSGLMNSQESH 960
QY 743 ----- 742
Db 961 GKNVSTESGRLFGKRAHPALLTKDNALFKVISILAKTKNTSNNATNKRTHIDGPSL 1020
QY 743 ----- 742
Db 1021 LIENSPVMQNLSESTEFKVKVTPLIHDRMLMDKNATALRLNHNMTSSKNMENVQOK 1080
QY 743 ----- 742
Db 1081 KEGPIPPAQNDPMSFFKMLPLPESARNIQRTHGKNSLNSGQSPKQLVSLGPEKSVKG 1140
QY 743 ----- 742
Db 1141 QNFLSEKNKVVVGKGFETKDVGLKEMVFPSSRNPLFLTLNLDLHNNTHNQEKIQEIEK 1200
QY 743 ----- 742
Db 1201 KETLIOENVVLPOIHTVTGKNFMKNLFLSTRQNVESYBAYAPVLQDPRSLNDSTNR 1260
QY 743 ----- 742
Db 1261 TKKHTAHSKGBEENLEGLGNQTKQIVEKYACTTRISPTNSQNFVQTSKRAKQFPL 1320
QY 743 ----- 742
Db 1321 PLEETELEKRIIVDDTSTQSKNMKHLTPSLTQIDYNEKEGAIQTQSPSLDCLTRSHSI 1380
QY 743 ----- 742
Db 1381 PQANRSPBIAKVSSPFRPIRYLTVLFDQNSHLPAAASYRKSGVQESSHPLQGAOK 1440
QY 743 ----- 742
Db 1441 NNLSLAILTEMTQDREVGLSTSATNSVTKKVENTVLPKDPLPKTSKVELLPKVHI 1500
QY 743 ----- 742
Db 1501 YOKDLFPPTETNSGPHGLDLVBSLLQTEGAIKWNEANRPGKVPFLRVATESSAKTFSK 1560
QY 743 ----- 742
Db 1561 LADPLANDNHGTYQIPKEENKQSKSPEKTAFAKKDITLSLNACSNHAIAINQGNKP 1620
QY 743 ----- 786
Db 1621 ELEVWAKQGRTERLCSQNPVLRKHQREITRTTLQSDQDEIDYDDTISVEMKEDFDIY 1680
QY 787 DEENQSPRSFQKTRHYFIRAVRLWDYGMSSSHVLNRAQSGSVQPKKVFQETD 846
Db 1681 DEENQSPRSFQKTRHYFIRAVRLWDYGMSSSHVLNRAQSGSVQPKKVFQETD 1740
QY 847 GSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEQORQA 906
Db 1741 GSFTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEQORQA 1800
QY 907 EPRKQVKNETKTYFWKVOHMAPTKDBFDCWAYFSDVDLEKDVHSGLIGPLLVCHT 966
Db 1801 EPRKQVKNETKTYFWKVOHMAPTKDBFDCWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
QY 967 NTLNPAHQGVQTVQSFALPFTTIDETKSWYFTFENNERCRAPCNTOMEDPTFKNYRHA 1026
Db 1861 NTLNPAHQGVQTVQSFALPFTTIDETKSWYFTFENNERCRAPCNTOMEDPTFKNYRHA 1920
QY 1027 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEBYKALNLYP 1086
Db 1921 INGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEBYKALNLYP 1980

RESULT 8
US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-8508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cdna sequence
```

US-08-212-133A-2

Query Match 94.1%; Score 7234; DB 1; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1439; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

```
QY 1 ATRRYYLGAVELSWDYNQSDLGELPVDARPPPPVPKSPFNTSVVYKTLFVEFTVHLFN 60
DB 1 ATRRYYLGAVELSWDYNQSDLGELPVDARPPPPVPKSPFNTSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPWGLGPTTQAEVYDVTWVITLKNMASHPVSLHAGVSYWKASGEGAYDDOTSQ 120
DB 61 IAKPRPPWGLGPTTQAEVYDVTWVITLKNMASHPVSLHAGVSYWKASGEGAYDDOTSQ 120
QY 121 REKEDDKVFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLMSGLIGALLVCR 180
DB 121 REKEDDKVFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLMSGLIGALLVCR 180
QY 181 EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDDAASARAWPKMHTVNGVNR 240
DB 181 EGSLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDDAASARAWPKMHTVNGVNR 240
QY 241 SLPLGLGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300
DB 241 SLPLGLGCHRKSVYMHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFLTAQTLL 300
QY 301 MDLGOFLLFCHISSHQHDGMEAVYKVDSCPEEPQLRMKNEEABDYDDLTDSEMDVVRP 360
DB 301 MDLGOFLLFCHISSHQHDGMEAVYKVDSCPEEPQLRMKNEEABDYDDLTDSEMDVVRP 360
QY 361 DDNSPSFQIRSAVKHPKWTWHYTAABEEDWDYAPLVLPDDRKYQSLNNGPQRIQ 420
DB 361 DDNSPSFQIRSAVKHPKWTWHYTAABEEDWDYAPLVLPDDRKYQSLNNGPQRIQ 420
QY 421 RYKVKRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIFKNQASRPNIYPHGI 480
DB 421 RYKVKRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTLIIFKNQASRPNIYPHGI 480
QY 481 TDVRPLYRRLPKGVGHLDKDFPILPGEIIFKYKWTVTVEDGPKSDPRCLTRYSSPVNME 540
DB 481 TDVRPLYRRLPKGVGHLDKDFPILPGEIIFKYKWTVTVEDGPKSDPRCLTRYSSPVNME 540
QY 541 RDLASGLIGPLLCYKESVDORGNQIMSKRNVILFSVFDENRSWLTENIORFLNPAG 600
DB 541 RDLASGLIGPLLCYKESVDORGNQIMSKRNVILFSVFDENRSWLTENIORFLNPAG 600
QY 601 VOLEDPEFOASNIMHSINGVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
DB 601 VOLEDPEFOASNIMHSINGVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
QY 661 KMVYEDTLFLPPSGETVFMENPGILWILGCHNSDFRNGMTALLKVVSCDKNTGDYYE 720
DB 661 KMVYEDTLFLPPSGETVFMENPGILWILGCHNSDFRNGMTALLKVVSCDKNTGDYYE 720
QY 721 DSYEDISAYLLSKNNAIEPRSF ----- 742
DB 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI PENDIEKTDPFHAHTPMP 780
QY 743 ----- 742
DB 781 KIQNVSSDLMLLRQSPTHGILSLDQEAKEYTFSDDPSPCAIDSNNSLSEMTHFRPQ 840
QY 743 ----- 742
DB 841 LHHSGDMVTPPSGLQRLNEKLGITTAATELKKLDFKVSSTNNLISTIPSNLAAGTDN 900
QY 743 ----- 742
DB 901 TSSLGPPMPVHYDSOLDTTLFGKKSPLTESGGPLSLEENNDKLLBSGLMNSQESSW 960
QY 743 ----- 742
DB 961 GKNVSTESGRLFKGRAHPALLTKDNALFKVSI SLAKNTKNTSNNATNRKTHIDGPSL 1020
```

```
QY 743 ----- 742
DB 1021 LIENSPSVQWQNTLESDETEPKKVTPLIHDRLMDKNATALLNHNMSKNTTSSKQMEMVQOK 1080
QY 743 ----- 742
DB 1081 KEGPIPPDAQNPDMSFFKMLFLPESARWIQRTGKNKSLNSGGQSPKQLVSLGPBKSVBG 1140
QY 743 ----- 742
DB 1141 QNFLSEKXNVVVGKGEFTKDVGLKEMVFPSSNLFITNLNLHNNTHNQEKIIOEBIEK 1200
QY 743 ----- 742
DB 1201 KETLIQENVVLPOIHTVTGTKFMKNLFLLSRQNVGSEYEGAYAPVLQDPRSLDSTNR 1260
QY 743 ----- 742
DB 1261 TKKHTAHFSKKGEBENLEGLGNQTKQIVEKYACTTRISPNTSQQNFVTQSRKRALKQFRL 1320
QY 743 ----- 742
DB 1321 PLEETELEKRIIVDDTSTQWSKMKHLTPSTILTQIDYNEKEKGAITQSPSLDCLFRSHI 1380
QY 743 ----- 742
DB 1381 PQANRSPPIAKVSSPSPSIRPIYLRVLFDQNSHLPAAASYRKKDSGVQESSHFLQGAKK 1440
QY 743 ----- 742
DB 1441 NNLSAILTLEMTGDQREVGLTGSATNSVTYKKYVENTVLPKPDLPKTSKGVLELPKVHI 1500
QY 743 ----- 742
DB 1501 YQKOLFPPTETSNNGSPGHLDEVGSLLOSTEGAIKNNEANRPGKVPFLAVATESSAKT98K 1560
QY 743 ----- 742
DB 1561 LIDPLAWDNHYGTQIPKEBWSQESPEKTAPKAKDITILSNACSNHAIAINEGQNKP 1620
QY 743 ----- 786
DB 1621 ELEVWAKQGRTERLCSQNPPLKHHQREIRTTTLQSDQEBIDYDDTISVENMKDEDFIY 1680
QY 787 DEDENQSPRSFQKTRHYFIAAVERLWDYGHSSSPHVLNRNAQSGSVQPKKVVQFEBTD 846
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGHSSSPHVLNRNAQSGSVQPKKVVQFEBTD 1740
QY 847 GSFTOPLVRGELNEHLGGLGPIYIRAEVDNIMVTFRNOASRPSPYSYSSLI SYEEDQROGA 906
DB 1741 GSFTOPLVRGELNEHLGGLGPIYIRAEVDNIMVTFRNOASRPSPYSYSSLI SYEEDQROGA 1800
QY 907 EPRKNFVKNETKTYFWKVQHMAPTKDEFCWAYEYSDVDLEKDVHSGLIGLPLVCHT 966
DB 1801 EPRKNFVKNETKTYFWKVQHMAPTKDEFCWAYEYSDVDLEKDVHSGLIGLPLVCHT 1860
QY 967 NTLNPAHQRTVQBFALFFTIIDETKSWYFTEENMERNCRAPCNIOBDDPTFKENYRPHA 1026
DB 1861 NTLNPAHQRTVQBFALFFTIIDETKSWYFTEENMERNCRAPCNIOBDDPTFKENYRPHA 1920
QY 1027 INGYIMDTLPGVMAQDQRIEWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1086
DB 1921 INGYIMDTLPGVMAQDQRIEWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 1087 GVFFTEVMLPKAGIWRVECLIGSHLHAGMSTFLVYSNKQCTPLGWSHGHIRFOITAS 1146
DB 1981 GVFFTEVMLPKAGIWRVECLIGSHLHAGMSTFLVYSNKQCTPLGWSHGHIRFOITAS 2040
QY 1147 GOYGOWAPKLARLHYSGINSINAWSTKBPFSWIKVDLLAPMIITHGKTQBARQKFSLSYISQ 1206
DB 2041 GOYGOWAPKLARLHYSGINSINAWSTKBPFSWIKVDLLAPMIITHGKTQBARQKFSLSYISQ 2100
```

QY 1207 FIIMSLDGKWKOTYRGNSGTGLMVFVGNVDSSGKIKHINFPPIIARIYIRLHPHYSIRS 1266
Db 2101 FIIMSLDGKWKOTYRGNSGTGLMVFVGNVDSSGKIKHINFPPIIARIYIRLHPHYSIRS 2160
QY 1267 TLMELMGCDLNSCSPLGHESKAI SDAQITASSYFTNMFATWSPSKARLHLQGRSNAMR 1326
Db 2161 TLMELMGCDLNSCSPLGHESKAI SDAQITASSYFTNMFATWSPSKARLHLQGRSNAMR 2220
QY 1327 PQVNNPKWLQVDFQKTMKVTVGTTOGVKSLLSMTVKEFLISSODGHQWTLFPQNGKV 1386
Db 2221 PQVNNPKWLQVDFQKTMKVTVGTTOGVKSLLSMTVKEFLISSODGHQWTLFPQNGKV 2280
QY 1387 KVFQGNQDSPTPVVNSLDPPLLRILYRIHPQSWHOIALRMEVLGCEAODLY 1438
Db 2281 KVFQGNQDSPTPVVNSLDPPLLRILYRIHPQSWHOIALRMEVLGCEAODLY 2332

RESULT 9

US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 5744446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver CDNA sequence
US-08-474-503-2

Query Match 94.1%; Score 7234; DB 1; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
Db 1 ATRRYILGAVELSWDMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEFTVHLFN 60
QY 61 IAKPRPPMGLGPTTQAEVDTVTITLKNMASHPVSLHAGVSVYWKASEGAEYDDQTSQ 120

Db 61 IAKPRPPMGLGPTTQAEVDTVTITLKNMASHPVSLHAGVSVYWKASEGAEYDDQTSQ 120
QY 121 RKEDDKVPGGASHVTVMQVLKENGPMASDPCLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 RKEDDKVPGGASHVTVMQVLKENGPMASDPCLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGS LAKEKTQTLLHKFTLLPAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240
Db 181 EGS LAKEKTQTLLHKFTLLPAVDFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVYNR 240
QY 241 SLFGLIGCHRKSVYWHVIGMGTTPVHSLFEGHTFLVRNHRQASLEISPIITFLTAQTL 300
Db 241 SLFGLIGCHRKSVYWHVIGMGTTPVHSLFEGHTFLVRNHRQASLEISPIITFLTAQTL 300
QY 301 MDLGGPFLFCHTSSHQHDCMEAYVKVDCSPPEPOLRMKNNEAEVDLDTSEMDVVPF 360
Db 301 MDLGGPFLFCHTSSHQHDCMEAYVKVDCSPPEPOLRMKNNEAEVDLDTSEMDVVPF 360
QY 361 DDNSPSTFIQIRSVAKGHPKTVWHVIAAEEEDWDYAPLVAPDDRYSKYQLNNGPQRTG 420
Db 361 DDNSPSTFIQIRSVAKGHPKTVWHVIAAEEEDWDYAPLVAPDDRYSKYQLNNGPQRTG 420
QY 421 RYKVKVRMAYTDETFKTRTAIQHESGILGPLLYGEVGTLLIIPKQASRPNTYPHGI 480
Db 421 RYKVKVRMAYTDETFKTRTAIQHESGILGPLLYGEVGTLLIIPKQASRPNTYPHGI 480
QY 481 TDVRLYSRRLPKGVKHLKDFILPGETFKYKMTVTVEGPTKSDPRCLTRYYSFVNME 540
Db 481 TDVRLYSRRLPKGVKHLKDFILPGETFKYKMTVTVEGPTKSDPRCLTRYYSFVNME 540
QY 541 RDLASGLIGPLICYKESVDQRNOIMSDKRNVLFSVFDENRSWLTENIQFLPNPAG 600
Db 541 RDLASGLIGPLICYKESVDQRNOIMSDKRNVLFSVFDENRSWLTENIQFLPNPAG 600
QY 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFFSGYTFKH 660
Db 601 VQLEDPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNRGNTALLKVSSCDKNTGDIYE 720
Db 661 KMVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFRNRGNTALLKVSSCDKNTGDIYE 720
QY 721 DSYEDISAYLLSKNNAIBRSPF----- 742
Db 721 DSYEDISAYLLSKNNAIBRSPF----- 742
QY 743 ----- 742
Db 743 ----- 742
QY 781 KIQNVSSDLLMLLRQSPTPHGLSLDLQAEKVFETSDPPSGAIDSNNSLSEMTFRPQ 840
Db 743 ----- 742
QY 841 LHSQDMVFTPEGSLQLRLNEKLGTAATATLKLDFKVSSTSNLIISTIPSDNLAAGTDN 900
Db 743 ----- 742
QY 901 TSSLGPPSPMPVHYDQLDTTLFGKSSPLTESGGPLSLSENNDSKLAESGLMNSQESSW 960
Db 743 ----- 742
QY 961 GKNVSTESGRLPFGKRAHGPALLTKONALFKVVISLLKTKNTKNTSNNATNRKTHIDGPSL 1020
Db 743 ----- 742
QY 1021 LIENSPSVQWQNILEDTEFFKVTPLIHDRMLMDKNATLRLNHNMTTSSKNMEMVOOK 1080
Db 743 ----- 742
QY 1081 KEGPIPPDAQNPMSFFKQFLPESARWIQTHGKNSLNSCGQSPKQVLSLGEPEKSV 1140
Db 743 ----- 742
QY 1141 QNFLSEKKNVVGKGEFTKDVGLKEMVFPSSRNFLTLNLDNLHNNTNHNQEKIQBEEK 1200

QY 743 ----- 742
Db 1201 KETLIQENVVLPOIHTVTGKTFWKNLFLSLTRQNVGSYEGAYAPVLQDERSLNDSTNR 1260
QY 743 ----- 742
Db 1261 TKKHTAHSKKGBEENLEGLNQTQIVEKIACCTTRISPTNSQNFVQSRKALKQFRL 1320
QY 743 ----- 742
Db 1321 PLSETELEKRIIYDDTSTQSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
QY 743 ----- 742
Db 1381 POANRSPLIATKVSFPISIRPIYILTRVLFDQNSSHLPAASRYKKDSGVQBSHFLQAKK 1440
QY 743 ----- 742
Db 1441 NNLSLAILTLEMTGDQREVSLGTSATNSYTKKVENTVLPKPDLPKTSKVBLLPKVHI 1500
QY 743 ----- 742
Db 1501 YOKCLPFTSTSGPHGLDLVEGSLQTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLAWDNHYGTQIPKEWKSOEKSPEKTAFFKKOTILSLNACESNHAIAINEGQNK 1620
QY 743 ----- 786
Db 1621 EIEVTWAKQRTRELCSQNPVLKQREITRTTLDQDEIDYDDTISVEMKKEPDIY 1680
QY 787 DEBENQSPRQKTRHYFTAAVERLMDYQMSPPHVLNRQAQSGVPQKVVQFQFTD 846
Db 1681 DEBENQSPRQKTRHYFTAAVERLMDYQMSPPHVLNRQAQSGVPQKVVQFQFTD 1740
QY 847 GSTOPLYRGELNEHGLGPIYRAEVEDNIMWTERNOASRPYSFSSLSISVEDQROGA 906
Db 1741 GSTOPLYRGELNEHGLGPIYRAEVEDNIMWTERNOASRPYSFSSLSISVEDQROGA 1800
QY 907 EPRKNFVKNETTYFWKHMAPTKDEDFCKAWAFSDVDLEKDVHSLGLIGLPLVCHT 966
Db 1801 EPRKNFVKNETTYFWKHMAPTKDEDFCKAWAFSDVDLEKDVHSLGLIGLPLVCHT 1860
QY 967 NTLNPAHGRQVTVQEFALFTIIFDETKSWFTENMERNCRAPCNIOKEDPTFKENYRFA 1026
Db 1861 NTLNPAHGRQVTVQEFALFTIIFDETKSWFTENMERNCRAPCNIOKEDPTFKENYRFA 1920
QY 1027 INGVIMDTPLGLVMAQDQRLRWYLLSMGNSNIHSHFSGHVFTVRKKEVKMALNYLP 1086
Db 1921 INGVIMDTPLGLVMAQDQRLRWYLLSMGNSNIHSHFSGHVFTVRKKEVKMALNYLP 1980
QY 1087 GVFETVEMLPKAGIWRVECLIGELHLAGMSTLFLVYSNKKQPTPLGMAHGHIRDQITAS 1146
Db 1981 GVFETVEMLPKAGIWRVECLIGELHLAGMSTLFLVYSNKKQPTPLGMAHGHIRDQITAS 2040
QY 1147 GOYGONAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMLIHGKTKQARQKPSLYISQ 1206
Db 2041 GOYGONAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMLIHGKTKQARQKPSLYISQ 2100
QY 1207 FIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGKIHNFNPPIIARYIRLHPHTHYSIRS 1266
Db 2101 FIIMYSLDGKWKQTYRGNSTGTLMVFFGNVDSSGKIHNFNPPIIARYIRLHPHTHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMLPGMESKAI SDAQITASSYFTNFPATWSPSKARHLQGRSNAR 1326
Db 2161 TLRMELMGCDLNSCMLPGMESKAI SDAQITASSYFTNFPATWSPSKARHLQGRSNAR 2220
QY 1327 PQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTSMYVKRFLISSODGHWTLFFQNGKV 1386
Db 2221 PQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTSMYVKRFLISSODGHWTLFFQNGKV 2280

QY 1387 KVFQGNQDSFTPVNSLDPPLTRYLRIRHQPQSWVHQIALRMEVILGCEAQDLY 1438
Db 2281 KVFQGNQDSFTPVNSLDPPLTRYLRIRHQPQSWVHQIALRMEVILGCEAQDLY 2332

RESULT 10

US-08-670-707A-2
; Sequence 2, Application US/08670707A
; Patent No. 5859204
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,707A
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-08-670-707A-2

Query Match 94.1%; Score 7234; DB 2; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYLGAVELSWDMYQSDLGELPVDARFPFRPKSPFFNTSVYKTKLFFVEFTVHLFN 60
Db 1 ATRRYLGAVELSWDMYQSDLGELPVDARFPFRPKSPFFNTSVYKTKLFFVEFTVHLFN 60
QY 61 TAKRPPPMGLIGPTIQAEVYDVTVTILKNMASHPVSLHVGVSFWKASEGAEYDDQTSQ 120
Db 61 TAKRPPPMGLIGPTIQAEVYDVTVTILKNMASHPVSLHVGVSFWKASEGAEYDDQTSQ 120
QY 121 REKEDDKVPFGSGSHYTVQVLKENGPMASDPCLTYSYLSHVDLVKDLNSGLIGALLVCR 180

Db 121 REKEDKVPFGSGHTYVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLMSGLIGALLVCR 180
QY 181 EGS LAKEKTOTLHKFTLLFAVDEGKSWHSEKNSLMQORDAASARAPKMTYVGYNR 240
Db 181 EGS LAKEKTOTLHKFTLLFAVDEGKSWHSEKNSLMQORDAASARAPKMTYVGYNR 240
QY 241 SLPGITGCHRKSVYVHVIWIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300
Db 241 SLPGITGCHRKSVYVHVIWIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300
QY 301 MDLGQFLPCHTSSHQHDMWAYVYKVDSPESPQLRMKNNEBAEDYDDDLTDSMDVWVF 360
Db 301 MDLGQFLPCHTSSHQHDMWAYVYKVDSPESPQLRMKNNEBAEDYDDDLTDSMDVWVF 360
QY 361 DDNSPFSFIQIRSVAKKHPKTVVHYIAAEEEDWDVAPLAPDDRSYKSOYLNNGPORIG 420
Db 361 DDNSPFSFIQIRSVAKKHPKTVVHYIAAEEEDWDVAPLAPDDRSYKSOYLNNGPORIG 420
QY 421 RYKVKVRMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNTYPHGI 480
Db 421 RYKVKVRMAYTDETFKTREAIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNTYPHGI 480
QY 481 TDVRLYRRLPKGVKHLKOFILPGEIFKYKWTVTVEDGPKSDPRCLTRYYSFVWME 540
Db 481 TDVRLYRRLPKGVKHLKOFILPGEIFKYKWTVTVEDGPKSDPRCLTRYYSFVWME 540
QY 541 RDLASGLIPLIICYKESVDQGNQIMSDKRVILFVSFDEKNSVLTENIORELPNPAG 600
Db 541 RDLASGLIPLIICYKESVDQGNQIMSDKRVILFVSFDEKNSVLTENIORELPNPAG 600
QY 601 VOLEDPFQASNIMHSINGVYVDSIQLSVCLHEVAYWYILSICAQTDFLSVFSGYTPKH 660
Db 601 VOLEDPFQASNIMHSINGVYVDSIQLSVCLHEVAYWYILSICAQTDFLSVFSGYTPKH 660
QY 661 KAVYEDTLTLPFSGETVPMSENPCIMILGCHNSDFRNGMTALLKVSSCDXQGTGYE 720
Db 661 KAVYEDTLTLPFSGETVPMSENPCIMILGCHNSDFRNGMTALLKVSSCDXQGTGYE 720
QY 721 DSYEDIAYILSKNNAIPRSP----- 742
Db 721 DSYEDIAYILSKNNAIPRSP----- 742
QY 743 ----- 742
Db 781 KIQVSSSLLMLLRQSPTHGLSLDLQEAKEYETFSDDPPSGAIDSNNSLSEMTFRPQ 840
QY 743 ----- 742
Db 841 LHSQDMVFTPESGLOLRINEKLGTTAATELKKLDFKVSSTSNLIISTIPSDNLAAGTDN 900
QY 743 ----- 742
Db 901 TSSLGPPMPVHYDSQDITLFGKSSPLITESGGPLSLSEBNDKSLLESGLMNSQESSW 960
QY 743 ----- 742
Db 961 GKNVSTESGRLFKGRAGHPALLTKDNALFKVISLLKTNKTSNNSATNKRKTHIDGPSL 1020
QY 743 ----- 742
Db 1021 LIENSPTSVMQILESDBTKVTPLLIHDRMLMDKQATALRLNHNMSKNTSSKNMFWQOK 1080
QY 743 ----- 742
Db 1081 KEGPIPPAQNDMSFFXWLFIPESARWIORTHGKNSLNSQGPSPKQVLSLGEKSVBG 1140
QY 743 ----- 742
Db 1141 QNFLSEKNVWVGGEFTKQVGLKEMWFPSSRNLF/TLNHLHNHNTNQBKIKQERIEK 1200
QY 743 ----- 742

Db 1201 KETLIQENWVLPQIHTVTGTRKMFKNMLFLLSTRQNVESYEGAYAPVLQDPESLNDSTR 1260
QY 743 ----- 742
Db 1261 TKKHTPAHPSKKBEENLEGLGNQTKQIIVEKYACTTRISPTSQNFVTFORSKRALQKPL 1320
QY 743 ----- 742
Db 1321 PLEETELESKR.IIVDDTSTQSKNMKHLTPSILTQIDYNEKEKGALITQSPSLDCLTRSHI 1380
QY 743 ----- 742
Db 1381 POANRSLPIAKVSSFPSPIRPYLTRVLFDQNSHLPAAASYRKXKDSGVQESSHFLQGA 1440
QY 743 ----- 742
Db 1441 NNLSAILTLEMTGQREVSGLSATNSVTYKVKVENTVLPKPLPKTSGKVELLPKVI 1500
QY 743 ----- 742
Db 1501 YQKDLFPTETSGSPCHLDLVEGSLLOQTEGAIKWNEARPKVPFLRVATESAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLAWNHYGTQIPKEEWSQEKSPKTAPEKKKDTILSLNACESNHAIAINSGONKP 1620
QY 743 ----- 742
Db 1621 BIEVTMAQKQTERLCSQNPVLRKHQBETRTTLOSQOERIDYDDTTISVEMKXEDFDIY 1680
QY 787 DEDENQSPRSQKTRHYFIAAVERLDYGNSSSPHVLNRNAQSGSVQPKKVVFOBT 846
Db 1681 DEDENQSPRSQKTRHYFIAAVERLDYGNSSSPHVLNRNAQSGSVQPKKVVFOBT 1740
QY 847 GSFTQPLVGBELNHLGLLGPYIRAEVEDNIMVTFRNOASPYFYSSLSIYEEDQOQA 906
Db 1741 GSFTQPLVGBELNHLGLLGPYIRAEVEDNIMVTFRNOASPYFYSSLSIYEEDQOQA 1800
QY 907 EPRKNFVKNETKYFKVQOHMAPTDEPDCKAWAYFSDVDLEKDVHSGILGILLVCHT 966
Db 1801 EPRKNFVKNETKYFKVQOHMAPTDEPDCKAWAYFSDVDLEKDVHSGILGILLVCHT 1860
QY 967 NTLNPAHGRQVTVQEFALFFTIIDETKSWYFENMERNCRAPCNIQMEDPTFKENYRPHA 1026
Db 1861 NTLNPAHGRQVTVQEFALFFTIIDETKSWYFENMERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1027 INGYIMDTLPGVMAQOQRIWYLLSMGSNNIHSIHFSGHVFTVRKKEEYKMALYNLYP 1086
Db 1921 INGYIMDTLPGVMAQOQRIWYLLSMGSNNIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 1087 GVFEVEMLPKAGIWRVECLIGSHLAGHAGSTLFLVYSNKCOTPLGNASGHIRDFQITAS 1146
Db 1981 GVFEVEMLPKAGIWRVECLIGSHLAGHAGSTLFLVYSNKCOTPLGNASGHIRDFQITAS 2040
QY 1147 GQYQMAAPKLARLHYSINAWSTKEPFSWKVOLLAPMIHGIKTQCARQKPSLSIYQ 1206
Db 2041 GQYQMAAPKLARLHYSINAWSTKEPFSWKVOLLAPMIHGIKTQCARQKPSLSIYQ 2100
QY 1207 FIIMYSLDGKKWQYRNGSTGTLMVFFGNVDSGIGKININPPIIARYIRLHPHYSIRS 1266
Db 2101 FIIMYSLDGKKWQYRNGSTGTLMVFFGNVDSGIGKININPPIIARYIRLHPHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMLPGMESKAIQITASSYFTNMETATWSPSKARLHLOGRSNWR 1326
Db 2161 TLRMELMGCDLNSCMLPGMESKAIQITASSYFTNMETATWSPSKARLHLOGRSNWR 2220
QY 1327 POVNNPKWELQVDFQKTMKVTVGTQGVKSLLTSMYVKEBELISSODQHOMTLFFQNGKV 1386
Db 2221 POVNNPKWELQVDFQKTMKVTVGTQGVKSLLTSMYVKEBELISSODQHOMTLFFQNGKV 2280
QY 1387 KVFQGNQDSTPVPVNSLDPPLLTRYLRIRHPQSWHQAIALAMEVLGCBAQDLY 1438
Db 2281 KVFQGNQDSTPVPVNSLDPPLLTRYLRIRHPQSWHQAIALAMEVLGCBAQDLY 2332

RESULT 11

US-09-037-601-2
Sequence 2, Application US/09037601
Patent No. 6180371
GENERAL INFORMATION:
APPLICANT: Lollax, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/037,601
FILING DATE: 26-JUN-1996

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:

NAME: Perber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Liver

US-09-037-601-2

Query Match 94.1%; Score 7234; DB 3; Length 2332;
Best Local Similarity 61.7%; Pred. NO. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY	1	ATRRYYLGAVELSWDMQSDLGELPVDARFPFRPKSPFPNTSVVYKKTFLVEFTVHLFN	60
DB	1	ATRRYYLGAVELSWDMQSDLGELPVDARFPFRPKSPFPNTSVVYKKTFLVEFTVHLFN	60
QY	61	IAKPRPPMGLLGPTIQAEVYDTVTITLKNMASHPVSLHAGVSVYWKASGAEYDDQTSQ	120
DB	61	IAKPRPPMGLLGPTIQAEVYDTVTITLKNMASHPVSLHAGVSVYWKASGAEYDDQTSQ	120
QY	121	REKEDDKVPPGSGSHYVMQVLKENGPMASDPICLTYSYLSHVLDKDLNSGLIGALLVCR	180
DB	121	REKEDDKVPPGSGSHYVMQVLKENGPMASDPICLTYSYLSHVLDKDLNSGLIGALLVCR	180

QY	181	EGSLAKERTQTLHKPFIILFAVFDGKSMHSEKSLMMDRDAASARAWPKMHTVNGYVNR	240
DB	181	EGSLAKERTQTLHKPFIILFAVFDGKSMHSEKSLMMDRDAASARAWPKMHTVNGYVNR	240
QY	241	SLPGLIGCHRSVYVHWVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTLL	300
DB	241	SLPGLIGCHRSVYVHWVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTLL	300
QY	301	MDLGQFLFCHISSHQHDGMEAYVKVDSCEPPEQLRMKNNEBEADYDDDLTDSMDVYVRF	360
DB	301	MDLGQFLFCHISSHQHDGMEAYVKVDSCEPPEQLRMKNNEBEADYDDDLTDSMDVYVRF	360
QY	361	DDNSPSFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLAAPDRSRYKQYLNNQPQIG	420
DB	361	DDNSPSFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLAAPDRSRYKQYLNNQPQIG	420
QY	421	RKYKKVREMYATDETFKTREAIQHSGLTGLPLLGEVGDITLLIFKNOASPYNIYPHGI	480
DB	421	RKYKKVREMYATDETFKTREAIQHSGLTGLPLLGEVGDITLLIFKNOASPYNIYPHGI	480
QY	481	TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTRYYSFVNME	540
DB	481	TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTRYYSFVNME	540
QY	541	RDLAGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSHWLTENTQRLPNPAG	600
DB	541	RDLAGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSHWLTENTQRLPNPAG	600
QY	601	VOLEDPFOASNIHMSINGVYVDFSLSQLSVCLHEVAYVILSIGACTDLSVFFSGYTPKH	660
DB	601	VOLEDPFOASNIHMSINGVYVDFSLSQLSVCLHEVAYVILSIGACTDLSVFFSGYTPKH	660
QY	661	KWYEDTILTLFPFSGETVFMSENFGLMWILCHNSDFRNQMTALLKVSSCDKNTGDIYE	720
DB	661	KWYEDTILTLFPFSGETVFMSENFGLMWILCHNSDFRNQMTALLKVSSCDKNTGDIYE	720
QY	721	DSYEDISAYLLSKNAIEPRSP	742
DB	721	DSYEDISAYLLSKNAIEPRSP	742
QY	743	-----	742
DB	743	-----	742
QY	781	KIQNVSSSDJMLLRQSPHPGLSLSDLOBAKYETFFSDPSPGAIDSNNSLSEMTFRPQ	840
DB	743	-----	742
QY	841	LHSGDMVFTPESGIQLRLNEKLGTTAATLKLDFKVSSTSNLIPTIPSDNLAAGTDN	900
DB	743	-----	742
QY	901	TSSIGPPSPMPVHDSQDITLFGKSSPLTESGGPLSLEENNDKLLIESGLMNSQESSW	960
DB	743	-----	742
QY	961	GKNVSTESGRLFKGKRAHGAPALLTKDVALFKVSIILKTKNTKNSATNRKTHIDGPSL	1020
DB	743	-----	742
QY	1021	LIENSPVWQNIILSDTEFKKVTPLIHDRMLMDKNATLRNLNHNKTTSSKNMEMVQOK	1080
DB	743	-----	742
QY	1081	KEGPIPPDAQNPDMSPFKMLFLPESARWIQRTKNSLNSGQSPSPKQVSLGPEKSVEG	1140
DB	743	-----	742
QY	1141	QNFLEKKNVVGKGEFTKQVGLKEMVFPSSRLFLTNLDLNHNTHNQEKKEIEK	1200
DB	743	-----	742
QY	1201	KETLIQENVVLQIHVTGTGKNFMKNLFLSTRQNVESGYEGYAPVLQDFRSLDSTNR	1260
DB	743	-----	742

QY 601 VOLEDEPFOASNIMESINGVYFDSLSQLSVCLHEVAYWYLSIGAQTDFLSVFFSGYTFKH 660
DB 601 VOLEDEPFOASNIMESINGVYFDSLSQLSVCLHEVAYWYLSIGAQTDFLSVFFSGYTFKH 660
QY 661 KMYEDTTLTLPFSGETVPMSENGLWILGCHNSDFRNGHTALLKUSSCDKNTGDIYE 720
DB 661 KMYEDTTLTLPFSGETVPMSENGLWILGCHNSDFRNGHTALLKUSSCDKNTGDIYE 720
QY 721 DSYEDISAYLLSKNAIEPRSE----- 742
DB 721 DSYEDISAYLLSKNAIEPRSE----- 742
QY 743 ----- 742
DB 781 KTONVSSDLLLRQSPTPHGLSLSDLOEAKYETFDSPSPGAISSNNLSSEMTFRPQ 840
QY 743 ----- 742
DB 841 LHSQGVFTPESGLQLRLNEKLGTTAATELKKLDKVSSTNNLSTIPSNLAAGTDN 900
QY 743 ----- 742
DB 901 TSSLGPSPMPHYVDQDITLTKKXSSPLTESGGPLSLSEENNDKLLSGLMNSQESSW 960
QY 743 ----- 742
DB 961 GRNVSTESGRFLKGRAGHPALLTKDNALFKVSISSLTKNTKTSNNSATNKRTHIDGPSL 1020
QY 743 ----- 742
DB 1021 LIENSFVQWNILESDBTEKKVTPLIHDRMLMDKNATRLNHNKNTSSKMEMVQOK 1080
QY 743 ----- 742
DB 1081 KBGPDPDAQNDPMGFFKMLFLPESARMTIQRTHGKNSLNGSQSPKQLVSLGPKSVBG 1140
QY 743 ----- 742
DB 1141 QNFLSEKKNVVGKEFTKDVGLKEMVPPSRNLFNLNLDNHNNTNHEKKIOEBIEK 1200
QY 743 ----- 742
DB 1201 KETLIQENVVLPIHTVTGKNFMKNLFLSTRQNVGSEYAGAYAPVLQDFRSLNDSNR 1260
QY 743 ----- 742
DB 1261 TKGHTAHFSKGEENLEGLGNQTKQIVEKYACTTRISPNTSQNFVTTQSRKALKQFRL 1320
QY 743 ----- 742
DB 1321 PLEETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEGALITQSPSLDCLTRSHI 1380
QY 743 ----- 742
DB 1381 POANRSPFIKAVSPFSIRPIYLTRVLFDNSSHLPAAVYRKDSGVQESSHFLQJAKK 1440
QY 743 ----- 742
DB 1441 NNLAILTLEMTGQOREVSGSLGTSATNSVTYKKVENTVLPKDLPTKSGKVELLPKVIH 1500
QY 743 ----- 742
DB 1501 YQKDLFPPTETSNGPSCHLDLVEGSLLOQTEGAIKNNEARPKVPFLRVATESSAKTPSK 1560
QY 743 ----- 742
DB 1561 LLDPLAWDNHYGTQIPKEEWKSOEKSPEKTAFAKKDTILSLNACESNHAIAINEGQNKP 1620
QY 743 ----- 786
DB 1621 ELEVWAKGRTERLCSQNPPLKQHQREITITLQSDQEEIDYDDTISVENMKEDFDIY 1680
QY 787 DSDENQSPRSFQKTRHYFAAVERLWDYGMSSSPHVLNRAQSGSVFPQKKVVFQETD 846

DB 1681 DEDENQSPRSFQKTRHYFAAVERLWDYGMSSSPHVLNRAQSGSVFPQKKVVFQETD 1740
QY 847 GSFTQPLVRGSLNEHLGLGPYIRAEVEDNIMVTRFOASRPSPYSSLSIYEBEDQOGA 906
DB 1741 GSFTQPLVRGSLNEHLGLGPYIRAEVEDNIMVTRFOASRPSPYSSLSIYEBEDQOGA 1800
QY 907 EPRKNFVKPNETKTYFMKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGSLIGLVLVCHT 966
DB 1801 EPRKNFVKPNETKTYFMKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGSLIGLVLVCHT 1860
QY 967 NTLPAHQROVTVQEPALFFTFIDETKSWYTEMERNCRAPCMIQMEDPTFKENYRPHA 1026
DB 1861 NTLPAHQROVTVQEPALFFTFIDETKSWYTEMERNCRAPCMIQMEDPTFKENYRPHA 1920
QY 1027 INGIMDTLPGLVMAQDORIRWYLLSMGSENHIIHFSGHVFTVRKKEEYKMALYNLYP 1086
DB 1921 INGIMDTLPGLVMAQDORIRWYLLSMGSENHIIHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 1087 GVFTVEMLPSKAGIRWVECLIGBHLHAGMSTLFLVYSNKKQOTPLGMAHGHIRDFQITAS 1146
DB 1981 GVFTVEMLPSKAGIRWVECLIGBHLHAGMSTLFLVYSNKKQOTPLGMAHGHIRDFQITAS 2040
QY 1147 GOYQWAPKLARLYHSGSINAWSTKEPFSWIKVOLLAPMIITHGIKTQGAROKFSLSYISQ 1206
DB 2041 GOYQWAPKLARLYHSGSINAWSTKEPFSWIKVOLLAPMIITHGIKTQGAROKFSLSYISQ 2100
QY 1207 FIIMYSLDGKKQWYRGNSGTLMVFGNVDSGKKNIPNPPIIARIYIRLHPHYSIRS 1266
DB 2101 FIIMYSLDGKKQWYRGNSGTLMVFGNVDSGKKNIPNPPIIARIYIRLHPHYSIRS 2160
QY 1267 TRLMELMGCDLNSCMLPLGMSKASIDAQITPASSYFTNMFTATSPSKARLHLQGRSNAWR 1326
DB 2161 TRLMELMGCDLNSCMLPLGMSKASIDAQITPASSYFTNMFTATSPSKARLHLQGRSNAWR 2220
QY 1327 PAVNPNKEMLOWDFQKTKMKTGVTTOGVKSLTSMYKEFLISSODGHQHTLFPONGKV 1386
DB 2221 PAVNPNKEMLOWDFQKTKMKTGVTTOGVKSLTSMYKEFLISSODGHQHTLFPONGKV 2280
QY 1387 KVFQGNQDSFTPVVNSLDDPPLLTLYLRIHPQSWHQAIRMEVLGCEAQDLY 1438
DB 2281 KVFQGNQDSFTPVVNSLDDPPLLTLYLRIHPQSWHQAIRMEVLGCEAQDLY 2332

RESULT 13

US-09-523-656-2
; Sequence 2, Application US/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-951
; CURRENT APPLICATION NUMBER: US/09/523,656
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670,707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PAT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match 94.1%; Score 7234; DB 4; Length 2332;
Best Local Similarity 61.7%; Pred No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYILGAVELSWDYMQSDIGELPVDARFPFRVPSKSPFNTSVVYKTLFVEFTVHLEN 60
DB 1 ATRRYILGAVELSWDYMQSDIGELPVDARFPFRVPSKSPFNTSVVYKTLFVEFTVHLEN 60

QY	61	IAKPRPPWGLGPTIOAEVDTWITLKNMASHPVSLHVGVSYSKASEGAEYDDQTSQ	120
Db			
QY	61	IAKPRPPWGLGPTIOAEVDTWITLKNMASHPVSLHVGVSYSKASEGAEYDDQTSQ	120
Db			
QY	121	REKEDDKVPPGSHSYVWQVLKENGPMASDPLCLATYSYLSHVDLVKDLNSGLIGALLVCR	180
Db			
QY	121	REKEDDKVPPGSHSYVWQVLKENGPMASDPLCLATYSYLSHVDLVKDLNSGLIGALLVCR	180
Db			
QY	181	EGSLAKEKQTLHKFTLLFAVDEGKSWHSEKNSLMODRDAASARAWPKMHTVNGYNR	240
Db			
QY	181	EGSLAKEKQTLHKFTLLFAVDEGKSWHSEKNSLMODRDAASARAWPKMHTVNGYNR	240
Db			
QY	241	SUPGLIGCHRSYVHVHVGWTTPEVHSIFLGHFTFLVRNHRQASLEISPIFLTAQTLL	300
Db			
QY	241	SUPGLIGCHRSYVHVHVGWTTPEVHSIFLGHFTFLVRNHRQASLEISPIFLTAQTLL	300
Db			
QY	301	MDLGQPLLCFCHSSHQHDGMEAYKVDSCPEBPQLRMKNNEBAEDYDDDLTDSMDVVRP	360
Db			
QY	301	MDLGQPLLCFCHSSHQHDGMEAYKVDSCPEBPQLRMKNNEBAEDYDDDLTDSMDVVRP	360
Db			
QY	361	DDNPSFQIRSVAKKHKPTWHYIAAEEEDWDVAPLVLA PDRSYKSOYLNGPQRTG	420
Db			
QY	361	DDNPSFQIRSVAKKHKPTWHYIAAEEEDWDVAPLVLA PDRSYKSOYLNGPQRTG	420
Db			
QY	421	RYKVKVRFMAYTDEFTKTRAIQHESGILGPLLYGEVGDTLAIIPKQASRPYNTYPHGI	480
Db			
QY	421	RYKVKVRFMAYTDEFTKTRAIQHESGILGPLLYGEVGDTLAIIPKQASRPYNTYPHGI	480
Db			
QY	481	TDVRELYSRRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME	540
Db			
QY	481	TDVRELYSRRLPKGVKHLKDFILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME	540
Db			
QY	541	RLASGLIGPLICYKESVDQGNQIMSDKRVILFVSFEDENRSWLTENIQRFLEPNAG	600
Db			
QY	541	RLASGLIGPLICYKESVDQGNQIMSDKRVILFVSFEDENRSWLTENIQRFLEPNAG	600
Db			
QY	601	VOLEDPFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSICAQDTDFLSVFPFGYTFKH	660
Db			
QY	601	VOLEDPFOASNMHSINGYVFDLSQLSVCLHEVAYWYILSICAQDTDFLSVFPFGYTFKH	660
Db			
QY	661	KAVYEDTLTLPFSGETVPMSENPGILWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE	720
Db			
QY	661	KAVYEDTLTLPFSGETVPMSENPGILWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE	720
Db			
QY	721	DSYEDISAYLLSKNNAIEPRSF-----	742
Db			
QY	721	DSYEDISAYLLSKNNAIEPRSF-----	742
Db			
QY	743	-----	742
Db			
QY	781	KIQWVSSDLLMLLRQSPPHGLSLSLDQEAKEYTFSDPSPGAIIDSNNSLSEMTFRPQ	840
Db			
QY	743	-----	742
Db			
QY	841	LHSGDMVTPPSGLQLRNLNEKLTATATLKKLDFKVSSTSNLIISTIPSDNLAAGTDN	900
Db			
QY	743	-----	742
Db			
QY	901	TSSLGPPSMPVHVDSQLDTTLFGKXSSPLTESGGPLSLSEBNDKSLLESGLMNSQESSW	960
Db			
QY	743	-----	742
Db			
QY	961	GKNVSSSTESGLFKGKRAHPALLTKDNALFKVISLAKTNKTSNNSATNRKTHIDGPSL	1020
Db			
QY	743	-----	742
Db			
QY	1021	LIENSFSVWQNTLESDBTEPKVYPLIHDRMLMDKNATALLRNMGNKTTSSKNMBWQOK	1080
Db			
QY	743	-----	742
Db			
QY	1081	KEGP1PPAQNDMSFFKMLFLPESARW1QRTHGKNSLNSGQSPKQIVLSLGPKEVSVEG	1140
Db			

QY	743	-----	742
Db			
QY	1141	QNFLESEKKVYVKGCEFTKDVGLKEMVFPSSNLEFLTNLDNLHNNTHNQEKKIQBIEK	1200
Db			
QY	743	-----	742
Db			
QY	1201	KETLIQENVVLPIQIHTVTGTKNFMKNLELLSTRQNVGSEYEGAYAPVLQDPRSLDSTNR	1260
Db			
QY	743	-----	742
Db			
QY	1261	TKKHTAHPKSKGBEENLEGLGNQTKQIVKQVACTTRISPNTSQOQNFVTVQSRKALUKPRL	1320
Db			
QY	743	-----	742
Db			
QY	1321	PLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI	1380
Db			
QY	743	-----	742
Db			
QY	1381	PQANRSPPIAKVSSFPSIRPIYLTRVLFDQNSHLPAAASYRKKDSGVQESSHPLQGAJK	1440
Db			
QY	743	-----	742
Db			
QY	1441	NNLSAILTLEMTGQREVGLSGTSATNSVTYKXVENTVLPKPDLPKTSKGVKELLPKVHI	1500
Db			
QY	743	-----	742
Db			
QY	1501	YQKDLPTTETSGNSPGHLDLVEGSLJQCTEGAIKWNEANRPCKVPFLRVATESSAKTPSK	1560
Db			
QY	743	-----	742
Db			
QY	1561	LIDPLANDHYGTQIPKEBWSQESKPEKTAFFKKDTTILSNACESNHAIAINEQONKP	1620
Db			
QY	743	-----	742
Db			
QY	1621	BIEVTWAKQGBTERLCNQNPVLPKXKHOREIRTTLSQDQBEIDYDDTTISVENMKKEDFDIY	1680
Db			
QY	787	DEDENQSPRSQKTRHYFIAAVERLDYGMSSSPHVLNRNAQSGSVQFKKVVQBEFTD	846
Db			
QY	1681	DEDENQSPRSQKTRHYFIAAVERLDYGMSSSPHVLNRNAQSGSVQFKKVVQBEFTD	1740
Db			
QY	847	GSFTQPLVGEINBHLGGLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEDQOQA	906
Db			
QY	1741	GSFTQPLVGEINBHLGGLGPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEDQOQA	1800
Db			
QY	907	EPRKNFVKNETKTYEYKVOHMAFTKDFEFCCKAWAYESVDLKDQVHSGILGILLVCHT	966
Db			
QY	1801	EPRKNFVKNETKTYEYKVOHMAFTKDFEFCCKAWAYESVDLKDQVHSGILGILLVCHT	1860
Db			
QY	967	NTLPAHGRQVTVQBEFALFFTIFDETBSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA	1026
Db			
QY	1861	NTLPAHGRQVTVQBEFALFFTIFDETBSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA	1920
Db			
QY	1027	INGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHPSGHVFTVRKKEEYKMALYNLYP	1086
Db			
QY	1921	INGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHPSGHVFTVRKKEEYKMALYNLYP	1980
Db			
QY	1087	GVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCQTPLGSMASGHIRDFOQTAS	1146
Db			
QY	1981	GVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCQTPLGSMASGHIRDFOQTAS	2040
Db			
QY	1147	GOYQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ	1206
Db			
QY	2041	GOYQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKFSLSYISQ	2100
Db			
QY	1207	FIIMYSLDGKKWQYTRNGSTGTLNVFNGVNDSSGINKINIPNPPIIARYIRLPHYTHYSIRS	1266
Db			
QY	2101	FIIMYSLDGKKWQYTRNGSTGTLNVFNGVNDSSGINKINIPNPPIIARYIRLPHYTHYSIRS	2160
Db			
QY	1267	TLRMELMGCDLNSCMPGLMESKAI SPAQITASSYFTNMFTWSPSKARLHLQGRSNAR	1326
Db			
QY	2161	TLRMELMGCDLNSCMPGLMESKAI SPAQITASSYFTNMFTWSPSKARLHLQGRSNAR	2220
Db			
QY	1327	PQVNNPKWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV	1386
Db			

Db 1261 TKKHTAHFKSGEENLEGLNQTKQIVKYEYACTTRISPTSQNFVTVQSRKALKQFRL 1320
QY 743 ----- 742
Db 1321 PLEETELEKRIIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEGAIQSPSLDCLTRSHSI 1380
QY 743 ----- 742
Db 1381 PQANRSPPIAKVSPSPSIRPIYLRVLFDQNSHLPAASRYKKSGVQESSHFLQGANCK 1440
QY 743 ----- 742
Db 1441 NNLSLAILTEMTGDQREVGLSGTSATNSVTYKKVENTVLPKPDLPKTSKGVELLPKVHI 1500
QY 743 ----- 742
Db 1501 YQKOLFPTTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTRSK 1560
QY 743 ----- 742
Db 1561 LLDPLANDHNYGTQIPKEBWKQESKPEKTAFFKKDITLSLNACSNHAIAINRGQNK 1620
QY 743 -----SONPPVLKRHQREITRTTQSDQEEIDYDDTISVENMKEDFDIY 786
Db 1621 ELEVWAKQGRTERLCSQNPVPLKRHQREITRTTQSDQEEIDYDDTISVENMKEDFDIY 1680
QY 787 DDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAQSGSVPOFKKVVQEBFTD 846
Db 1681 DDENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLNRAQSGSVPOFKKVVQEBFTD 1740
QY 847 GSFTQPLYGELEHGLGCPYIRAEVEDNIMVTFNRAQSRPVSFYSSLSIYEEBQROCA 906
Db 1741 GSFTQPLYGELEHGLGCPYIRAEVEDNIMVTFNRAQSRPVSFYSSLSIYEEBQROCA 1800
QY 907 EPRKQFVKNETKTVFWKQVHMAPTKDEFCCKAWAYSDVDLEKDVHSGGLGILLVCHT 966
Db 1801 EPRKQFVKNETKTVFWKQVHMAPTKDEFCCKAWAYSDVDLEKDVHSGGLGILLVCHT 1860
QY 967 NTLNPAHQGVTVQEFALFTTIDETKSWYFTFENMERNCRAPCNQIMEDPTPKENYRTHA 1026
Db 1861 NTLNPAHQGVTVQEFALFTTIDETKSWYFTFENMERNCRAPCNQIMEDPTPKENYRTHA 1920
QY 1027 INGYIMDTPLGLVMAQDQIRWYLLSMGSGNENIHSHFSGHVTFTVRKKEEYKWAYLYNP 1086
Db 1921 INGYIMDTPLGLVMAQDQIRWYLLSMGSGNENIHSHFSGHVTFTVRKKEEYKWAYLYNP 1980
QY 1087 GVFETVEMLPKAGIWRVCECLIGEHLHAGMSTFLVYSNKQCTPLGMAHGHIRDQITAS 1146
Db 1981 GVFETVEMLPKAGIWRVCECLIGEHLHAGMSTFLVYSNKQCTPLGMAHGHIRDQITAS 2040
QY 1147 GYGOWAPKLARLHYSGSNASTKEPPSWIKVDLLAPMIHGKTKQAROKFESSLYISO 1206
Db 2041 GYGOWAPKLARLHYSGSNASTKEPPSWIKVDLLAPMIHGKTKQAROKFESSLYISO 2100
QY 1207 FIIMYSLDGKKWQTYRGNSGTGLMVFPGNVDSSGKKNIFNPPILARVIRLHPHYSIRS 1266
Db 2101 FIIMYSLDGKKWQTYRGNSGTGLMVFPGNVDSSGKKNIFNPPILARVIRLHPHYSIRS 2160
QY 1267 TLRMELMGCDLNSCMPLGWESKALSDAQITASSYFTNMFAWSPSKARLHLQGRSNAR 1326
Db 2161 TLRMELMGCDLNSCMPLGWESKALSDAQITASSYFTNMFAWSPSKARLHLQGRSNAR 2220
QY 1327 PQVNNPKWLOVDFOKTMKVGTGTVGVKSLATSMYVKEFLSSQDGHQWTLFPQNGKV 1386
Db 2221 PQVNNPKWLOVDFOKTMKVGTGTVGVKSLATSMYVKEFLSSQDGHQWTLFPQNGKV 2280
QY 1387 KVFQGNQDSFTPVNSLOPPLITRYLRIRHPQSWHVOIALRMEVLGCEAQDLY 1438
Db 2281 KVFQGNQDSFTPVNSLOPPLITRYLRIRHPQSWHVOIALRMEVLGCEAQDLY 2332

RESULT 15

PCT-US94-13200-2

; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patricia L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US94-13200-2

Query Match 94.1%; Score 7234; DB 5; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;
QY 1 ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLPN 60
Db 1 ATRRYLGAVELSWDYMOSDLGELPVDARPPRPVKSPFNTSVYKTLFVEFTVHLPN 60
QY 61 IAKPRPPMGLGPTIQAEVYDVTVTILKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ 120
Db 61 IAKPRPPMGLGPTIQAEVYDVTVTILKNMASHPVSLHVGVSVMKASEGAEYDDQTSQ 120
QY 121 REKEDDKVPFGSGTYYWQVLKENGPMASDPLCLTYSLSHVDLVKOLNSGLIGALLVCR 180
Db 121 REKEDDKVPFGSGTYYWQVLKENGPMASDPLCLTYSLSHVDLVKOLNSGLIGALLVCR 180
QY 181 EGSLLAKETQTLHAFILLFAVDFDGSKSWHSTKSLMDQDRDAASARAPKMHVYVNR 240
Db 181 EGSLLAKETQTLHAFILLFAVDFDGSKSWHSTKSLMDQDRDAASARAPKMHVYVNR 240
QY 241 SLPLGLICHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPTITLTAQTLL 300
Db 241 SLPLGLICHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSISPTITLTAQTLL 300
QY 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEEPOLRMKNNEEADYDDDLTDSMDVVR 360
Db 301 MDLQGFLLFCHISSHQHDGMEAYVKVDSCEEPOLRMKNNEEADYDDDLTDSMDVVR 360
QY 361 DDDNSPSFIQIRSVAKKHPTWVHYIAAEBEDMDYAPLVLPADDRSYKSYQLNNGPQRIG 420

Db	361	DDNSPSTQIRSAKHPKTYWYIAAEEDWDYAFVLAPDDRSYKSYQLNNGPORIG	420
Qy	421	RKYKVRFMAYTDEFTKREAIQHESGILGPLYGEGVDTLLIIFKNOASRPYNIYPHGI	480
Db	421	RKYKVRFMAYTDEFTKREAIQHESGILGPLYGEGVDTLLIIFKNOASRPYNIYPHGI	480
Qy	481	TDVPLYSRRLLPKGVKHLKDPPIIPGBIFKYKWTVTVEDGPTKSDPCLTRYSSFVNME	540
Db	481	TDVPLYSRRLLPKGVKHLKDPPIIPGBIFKYKWTVTVEDGPTKSDPCLTRYSSFVNME	540
Qy	541	RDLASGLIGPLIICVKESVDORGNOIMSDXENVILFSVFDENRSWYLTENIQRLPNPAG	600
Db	541	RDLASGLIGPLIICVKESVDORGNOIMSDXENVILFSVFDENRSWYLTENIQRLPNPAG	600
Qy	601	VOLEDPEFQASINMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660
Db	601	VOLEDPEFQASINMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660
Qy	661	KWYVEDTLTLPFPFSGETVFKSMENPGLWILCHNSDFRNRMGTALLKVSSCDKXMTGYDE	720
Db	661	KWYVEDTLTLPFPFSGETVFKSMENPGLWILCHNSDFRNRMGTALLKVSSCDKXMTGYDE	720
Qy	721	DSYEDIGAYLLSKNAIEPRSF-----	742
Db	721	DSYEDIGAYLLSKNAIEPRSFQXSRHPSTRQKFNATTIPENDIEKTDPFHAHRTFMP	780
Qy	743	-----	742
Db	781	KIQNVSSDLLMLLRQSPTRHGLSLDLQKAYETFPDDPSGAI DNSNLSISEMTHFRPQ	840
Qy	743	-----	742
Db	841	LHHSQDMVFTPESGQLURLNEKLTGAATELKKLDFKVSSTSNMLISTIPSDNLAAGTDN	900
Qy	743	-----	742
Db	901	TSSLSGPMVPHYDSQLDYYTLFGKKSPLTESGGPLSLEENDSKLLEGLMNSQESSW	960
Qy	743	-----	742
Db	961	GKNVSTESGRLFKGKRAHGPAALLTKDNALFKVSI SLKTKNTKSNSATNRKTHIDGPSL	1020
Qy	743	-----	742
Db	1021	LIENSPVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRLNHSNKTTSSKNMEMVOOK	1080
Qy	743	-----	742
Db	1081	KEGIPPAQNPOMSPFKMLFLPESABWQRTHCXNSLNSGGQSPKQLVSLGPEKSVEG	1140
Qy	743	-----	742
Db	1141	QNFELSEKNVVGKGETKDVGLKEMVFPSSRNFLTLDNLHENHNTNQEKKIQEBIEK	1200
Qy	743	-----	742
Db	1201	KETLIQENVVLPOIHTVTGTNFKWNLFLLLSTRQNVESYGAYAPVLQDFRSLNDSTNR	1260
Qy	743	-----	742
Db	1261	TKKHTAFSKKBEENLEGLGNQTKQIVEKYACTTRISPTWTSQQNFVTOQRKALQKPL	1320
Qy	743	-----	742
Db	1321	PLEETELEKRIIVDDTSTQWSKNMKHLTPSTLQTIDYNEKEKGAIQSPSLDCLTRSHSI	1380
Qy	743	-----	742
Db	1381	PQANRSPLPITAKVSSPFSIPPIYLRVLFDQNSSHLPAAASYRKDSGVQESSHFLQKAKK	1440
Qy	743	-----	742

Db	1441	NNLSAILLTLEMTGDQREVSGSLGTSATNSVYKCKVENTVLPKPDLPKTSKVELLPKVHI	1500
Qy	743	-----	742
Db	1501	YOKDLPTETSNPGSHLDLVEGSLLOCTEGAIKWNEANRPCKVPFLRVATESSAKTPSK	1560
Qy	743	-----	742
Db	1561	LLDPLADNHYGTQIPKEBWKSOBKPEKTAFFKKDITILSNACESHAAINEGONKP	1620
Qy	743	-----	786
Db	1621	EIEVTHAQGBTELCSQNPVLKRLQRELTRTLOSDBEIDYDITISVENKKEDFDIY	1680
Qy	787	DEDENQSPRSFKKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVPOKKVQFEFTD	846
Db	1681	DEDENQSPRSFKKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVPOKKVQFEFTD	1740
Qy	847	GSTOPLVYGEINEHGLGLGPIYIRASVEDNIMVTFNRQASRPYSFYSSILSIYEEBORQA	906
Db	1741	GSTOPLVYRGLNEHGLGLGPIYIRASVEDNIMVTFNRQASRPYSFYSSILSIYEEBORQA	1800
Qy	907	EPRKNFVKPNETKTYFWKVQHMAPTKDEDFCKAMAYFSDVLEKDVHSGLLGPLLVCHT	966
Db	1801	EPRKNFVKPNETKTYFWKVQHMAPTKDEDFCKAMAYFSDVLEKDVHSGLLGPLLVCHT	1860
Qy	967	NTLNPAHGRQVTVQEBALPFTTIDETKSYFTFENMERNCRAPCNIQMEDPTPKENYRPHA	1026
Db	1861	NTLNPAHGRQVTVQEBALPFTTIDETKSYFTFENMERNCRAPCNIQMEDPTPKENYRPHA	1920
Qy	1027	INGVIMDTLPLVMAODQRIRWYLLSMGSGNENIHSIHFGSHVFTVRKKEEYKWAYLYNP	1086
Db	1921	INGVIMDTLPLVMAODQRIRWYLLSMGSGNENIHSIHFGSHVFTVRKKEEYKWAYLYNP	1980
Qy	1087	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKKOTPLGASGHIRDEQITAS	1146
Db	1981	GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKKOTPLGASGHIRDEQITAS	2040
Qy	1147	GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDDLAPMIHGIKTQGNRQKFSSLYISQ	1206
Db	2041	GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDDLAPMIHGIKTQGNRQKFSSLYISQ	2100
Qy	1207	FIIMYSLDGKKQVTVRGNSGTGLVWFFGNVDSGGIKHINFNPPIIARYIRLHPHYSIRS	1266
Db	2101	FIIMYSLDGKKQVTVRGNSGTGLVWFFGNVDSGGIKHINFNPPIIARYIRLHPHYSIRS	2160
Qy	1267	TLRMBELMGCDLNSCSMPLGWESKAISDAQITASSYFTNMFAFWSSPSKARLHLQGSNAWR	1326
Db	2161	TLRMBELMGCDLNSCSMPLGWESKAISDAQITASSYFTNMFAFWSSPSKARLHLQGSNAWR	2220
Qy	1327	PQVNNPKEWLQVDFQKTMKVTVGTQGVKSILTSVMYKFEFLISSQDGHQWTLFPQNGKV	1386
Db	2221	PQVNNPKEWLQVDFQKTMKVTVGTQGVKSILTSVMYKFEFLISSQDGHQWTLFPQNGKV	2280
Qy	1387	KVFOGHQDSTFTPVNSLDPLLTIRLIRHPQSGWHQIALRMEVLGCEAQDIY	1438
Db	2281	KVFOGHQDSTFTPVNSLDPLLTIRLIRHPQSGWHQIALRMEVLGCEAQDIY	2332

Search completed: December 9, 2003, 17:10:34
Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 17:03:15 ; Search time 43 Seconds
(without alignments)
6219.640 Million cell updates/sec

Title: US-10-047-257-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSMDYMSQD.....VWHQIALRVEVLGCEAQLY 1438

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

```

1: /cgn2_6/ptodata/1/pubpaa/PCT NEW PUB.pdb:
2: /cgn2_6/ptodata/1/pubpaa/PCT NEW PUB.pdb:
3: /cgn2_6/ptodata/1/pubpaa/US06 NEW PUB.pdb:
4: /cgn2_6/ptodata/1/pubpaa/US06 PUBCOMB.pdb:
5: /cgn2_6/ptodata/1/pubpaa/US07 NEW PUB.pdb:
6: /cgn2_6/ptodata/1/pubpaa/PCTUS PUBCOMB.pdb:
7: /cgn2_6/ptodata/1/pubpaa/US08 NEW PUB.pdb:
8: /cgn2_6/ptodata/1/pubpaa/US08 PUBCOMB.pdb:
9: /cgn2_6/ptodata/1/pubpaa/US09A PUBCOMB.pdb:
10: /cgn2_6/ptodata/1/pubpaa/US09B PUBCOMB.pdb:
11: /cgn2_6/ptodata/1/pubpaa/US09C PUBCOMB.pdb:
12: /cgn2_6/ptodata/1/pubpaa/US09 NEW PUB.pdb:
13: /cgn2_6/ptodata/1/pubpaa/US10A PUBCOMB.pdb:
14: /cgn2_6/ptodata/1/pubpaa/US10B PUBCOMB.pdb:
15: /cgn2_6/ptodata/1/pubpaa/US10C PUBCOMB.pdb:
16: /cgn2_6/ptodata/1/pubpaa/US10 NEW PUB.pdb:
17: /cgn2_6/ptodata/1/pubpaa/US60 NEW PUB.pdb:
18: /cgn2_6/ptodata/1/pubpaa/US60 PUBCOMB.pdb:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7691	100.0	1438	14	US-10-006-091-1
2	7691	100.0	1438	14	US-10-047-257-1
3	7691	100.0	1438	15	US-10-225-900-1
4	7674	99.8	1471	14	US-10-095-718-2
5	7234	94.1	2332	12	US-09-957-641-2
6	7234	94.1	2332	12	US-10-131-510A-2
7	7234	94.1	2332	12	US-10-187-319-2
8	7227	94.0	2351	12	US-10-133-907-4
9	7227	94.0	2351	12	US-10-132-829-4
10	7227	94.0	2351	15	US-10-172-712-27
11	6599	85.8	1431	14	US-10-095-718-4
12	6489	84.4	1443	12	US-10-131-510A-39
13	6489	84.4	1443	15	US-10-187-319-39
14	6227	81.0	2133	12	US-10-131-510A-37
15	6227	81.0	2133	15	US-10-187-319-37

```

16 6192 80.5 2319 12 US-10-131-510A-6
17 6192 80.5 2319 15 US-10-187-319-6
18 2389.5 31.1 2224 15 US-10-115-563-14
19 2389.5 31.1 2224 15 US-10-172-712-31
20 1742.5 22.7 1160 12 US-10-137-870-234
21 1742.5 22.7 1160 12 US-10-140-018-234
22 1742.5 22.7 1160 12 US-10-140-021-234
23 1742.5 22.7 1160 12 US-10-140-274-234
24 1742.5 22.7 1160 12 US-10-140-471-234
25 1742.5 22.7 1160 12 US-10-140-807-234
26 1742.5 22.7 1160 12 US-10-140-922-234
27 1742.5 22.7 1160 12 US-10-140-924-234
28 1742.5 22.7 1160 12 US-10-140-926-234
29 1742.5 22.7 1160 12 US-10-141-698-234
30 1742.5 22.7 1160 12 US-10-141-702-234
31 1742.5 22.7 1160 12 US-10-141-704-234
32 1742.5 22.7 1160 12 US-10-142-421-234
33 1742.5 22.7 1160 12 US-10-142-432-234
34 1742.5 22.7 1160 12 US-10-142-767-234
35 1742.5 22.7 1160 12 US-10-143-033-234
36 1742.5 22.7 1160 12 US-10-144-994-234
37 1742.5 22.7 1160 12 US-10-145-628-234
38 1742.5 22.7 1160 12 US-10-145-631-234
39 1742.5 22.7 1160 12 US-10-145-633-234
40 1742.5 22.7 1160 12 US-10-145-746-234
41 1742.5 22.7 1160 12 US-10-145-748-234
42 1742.5 22.7 1160 12 US-10-145-823-234
43 1742.5 22.7 1160 12 US-10-145-826-234
44 1742.5 22.7 1160 12 US-10-145-870-234
45 1742.5 22.7 1160 12 US-10-145-876-234

```

ALIGNMENTS

RESULT 1

```

US-10-006-091-1
; Sequence 1, Application US/10006091
; Publication No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006.091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-006-091-1

```

Query Match 100.0%; Score 7691; DB 14; Length 1438;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATRRYLGAVELSMDYMSQDGLPVDARPPRPVKSPFNTSVYKKTLFVFTVHLFN 60
Db 1 ATRRYLGAVELSMDYMSQDGLPVDARPPRPVKSPFNTSVYKKTLFVFTVHLFN 60
Qy 61 IAKPRPFWGLLFTTQAEVYDVTWTLKMAHSPVLSHAGVSYNKAISGAEDDOTSQ 120
Db 61 IAKPRPFWGLLFTTQAEVYDVTWTLKMAHSPVLSHAGVSYNKAISGAEDDOTSQ 120
Qy 121 REKEDKVPFGSGHTYVQVYLKENGPMASDPLCLITYSLYSHVDLVKDLNSGLICALIVCR 180

```

Db 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGS LAKEKTOTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 181 EGS LAKEKTOTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLFGLGCHRSKYVWVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Db 241 SLFGLGCHRSKYVWVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTLL 300
QY 301 MDLGOQLLFCHISHHQHDGMEAVKVDSCPEEPQLRMKNNEBAEDVDDDLTDSMDVVRP 360
Db 301 MDLGOQLLFCHISHHQHDGMEAVKVDSCPEEPQLRMKNNEBAEDVDDDLTDSMDVVRP 360
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVAPDDRYSKYQLNNGPQIRG 420
Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVAPDDRYSKYQLNNGPQIRG 420
QY 421 RYKXKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGTDLIIIFKNQASRPNIYPHGI 480
Db 421 RYKXKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGTDLIIIFKNQASRPNIYPHGI 480
QY 481 TDVRPLYSRRLPGVXHLKDFPLPGELFKYKWTVTVEGPTKSDPRCLTRYYSFVWME 540
Db 481 TDVRPLYSRRLPGVXHLKDFPLPGELFKYKWTVTVEGPTKSDPRCLTRYYSFVWME 540
QY 541 RDLASGLIGPLLI CYKESVDQRGNQIMSKRNVLFSVFDENKSWLTENIQRFPLNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQRGNQIMSKRNVLFSVFDENKSWLTENIQRFPLNPAG 600
QY 601 VQLEDEPFOASNIWHSINGVVPDLSQLSVCLHEVAYWYLLISGAQTDLSVFPSTPKH 660
Db 601 VQLEDEPFOASNIWHSINGVVPDLSQLSVCLHEVAYWYLLISGAQTDLSVFPSTPKH 660
QY 661 KMVYEDTLTFPFGSETVPMENPGLWILGCHNSDFRNGMTALLKVSCKDKNTDYIE 720
Db 661 KMVYEDTLTFPFGSETVPMENPGLWILGCHNSDFRNGMTALLKVSCKDKNTDYIE 720
QY 721 DSYEDISAYLLSKNNAIEPRSPONPPVLKRGHOREITRTTLOSDQBEIDYDITISVEMKK 780
Db 721 DSYEDISAYLLSKNNAIEPRSPONPPVLKRGHOREITRTTLOSDQBEIDYDITISVEMKK 780
QY 781 EDFDIYDEDNQSPRFOKTRHYFJAVERLWDYGMSSSPHVLNRRAQSGVPQPKVY 840
Db 781 EDFDIYDEDNQSPRFOKTRHYFJAVERLWDYGMSSSPHVLNRRAQSGVPQPKVY 840
QY 841 FQEFDTGSGFTQPLYRGELNEHLGLGFPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900
Db 841 FQEFDTGSGFTQPLYRGELNEHLGLGFPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 900
QY 901 DQROGAEPKRNFKPNETKTYFKVQVHNAPTDEDFCKAWAYFSDVDLEKDVHSLGIGP 960
Db 901 DQROGAEPKRNFKPNETKTYFKVQVHNAPTDEDFCKAWAYFSDVDLEKDVHSLGIGP 960
QY 961 LLVCHTNTLNPAGRQVTVQEFALFTTIFDETSKSWYFTENMERNCRAPCNQIMEDTPEKE 1020
Db 961 LLVCHTNTLNPAGRQVTVQEFALFTTIFDETSKSWYFTENMERNCRAPCNQIMEDTPEKE 1020
QY 1021 NYRFHAINGYIMDTLGLVMAQQRIRWYLLSGNSNENIHSIFHSGHVTFRKKEFYKA 1080
Db 1021 NYRFHAINGYIMDTLGLVMAQQRIRWYLLSGNSNENIHSIFHSGHVTFRKKEFYKA 1080
QY 1081 LYNLYPGVETVEMLPKAGIWEVCLIGHLHAGNSTLPLVYSNCKOTPLGWSGHID 1140
Db 1081 LYNLYPGVETVEMLPKAGIWEVCLIGHLHAGNSTLPLVYSNCKOTPLGWSGHID 1140
QY 1141 FQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGKIQGARQKES 1200
Db 1141 FQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGKIQGARQKES 1200
QY 1201 SLYISOFIIMYSIDGKKWQYIRGNSTGTLWVFGNVDSSGICKHINFPPIIARYIRLHPT 1260
Db 1201 SLYISOFIIMYSIDGKKWQYIRGNSTGTLWVFGNVDSSGICKHINFPPIIARYIRLHPT 1260

RESULT 2

US-10-047-257-1

; Sequence 1, Application US/10047257

; Publication NO. US20020115152A1

; GENERAL INFORMATION:

; APPLICANT: Cho, Myung-Sam

; APPLICANT: Chan, Sham-Yuen

; APPLICANT: Kelsey, William

; APPLICANT: Yee, Helena

; TITLE OF INVENTION: Expression System for Factor VIII

; FILE REFERENCE: MSB-7255.2

; CURRENT APPLICATION NUMBER: US/10/047,257

; CURRENT FILING DATE: 2002-01-15

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1438

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Derived from

; OTHER INFORMATION: human factor VIII sequence

US-10-047-257-1

Query Match 100.0%; Score 7691; DB 14; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATRRYTLGAVELSWNDYMSDGLGELPVDARFPFRVPKSPFPNTSVVYKKTFLFVEFTVHLFN 60
Db 1 ATRRYTLGAVELSWNDYMSDGLGELPVDARFPFRVPKSPFPNTSVVYKKTFLFVEFTVHLFN 60
QY 61 IAKPRPPMGLLGPITQAEVYDVTVI TLKNMASHPVSLHVGVS YWKASBEGAYDDQTSQ 120
Db 61 IAKPRPPMGLLGPITQAEVYDVTVI TLKNMASHPVSLHVGVS YWKASBEGAYDDQTSQ 120
QY 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
Db 121 REKEDDKVFPGGSHYVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 181 EGS LAKEKTOTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
Db 181 EGS LAKEKTOTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLFGLGCHRSKYVWVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTLL 300
Db 241 SLFGLGCHRSKYVWVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIFTLTAQTLL 300
QY 301 MDLGOQLLFCHISHHQHDGMEAVKVDSCPEEPQLRMKNNEBAEDVDDDLTDSMDVVRP 360
Db 301 MDLGOQLLFCHISHHQHDGMEAVKVDSCPEEPQLRMKNNEBAEDVDDDLTDSMDVVRP 360
QY 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVAPDDRYSKYQLNNGPQIRG 420
Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVAPDDRYSKYQLNNGPQIRG 420
QY 421 RYKXKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGTDLIIIFKNQASRPNIYPHGI 480
Db 421 RYKXKVRPMAYTDETFKTRAIQHESGILGPLLYGEVGTDLIIIFKNQASRPNIYPHGI 480

481 TDVRLYRRLPKGVKHLKDFPILRGIFPKYKMTVTVEDGPTKSDPRCLTRYSSFNME 540
481 TDVRLYRRLPKGVKHLKDFPILRGIFPKYKMTVTVEDGPTKSDPRCLTRYSSFNME 540
541 RDLASGLIGPLLICYKESVDQRGNQMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
541 RDLASGLIGPLLICYKESVDQRGNQMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
601 VQLEDPFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
601 VQLEDPFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
661 KMVYEDTLTLFPFSGETVMSMENPGLWILCHNSDPRNGMTALLKVSSCDKNTGDIYE 720
661 KMVYEDTLTLFPFSGETVMSMENPGLWILCHNSDPRNGMTALLKVSSCDKNTGDIYE 720
721 DSYEDISAYLLSKNNAIEPRSFQKTRHYFIAAVERLDYGNSSPHVLRNRAQSGSVQPKVY 840
721 DSYEDISAYLLSKNNAIEPRSFQKTRHYFIAAVERLDYGNSSPHVLRNRAQSGSVQPKVY 840
841 FQFTDGSFTQPLRYGELNEHLGLGPGYIRAEVEDNIMVFRNOAGRPYFYSLSIYER 900
841 FQFTDGSFTQPLRYGELNEHLGLGPGYIRAEVEDNIMVFRNOAGRPYFYSLSIYER 900
901 DQCGABPRKFNKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDVLEKDVHSLGIP 960
901 DQCGABPRKFNKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDVLEKDVHSLGIP 960
961 LLVCHTNLPAHGRQVTVQEFALFTTIDETKSWFTENMERNCRAPCNIMEDTPEKE 1020
961 LLVCHTNLPAHGRQVTVQEFALFTTIDETKSWFTENMERNCRAPCNIMEDTPEKE 1020
1021 NYRPHAINGYIMDTLPLGVAQDQIRWYLLSGNSNENIHSIFSGHVFTVRKKEEYKMA 1080
1021 NYRPHAINGYIMDTLPLGVAQDQIRWYLLSGNSNENIHSIFSGHVFTVRKKEEYKMA 1080
1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGHSTLFLVYNSKQCTPLGMAHGIRD 1140
1081 LYNLYPGVFTVEMLPKAGIWRVECLIGELHAGHSTLFLVYNSKQCTPLGMAHGIRD 1140
1141 FQITASGOYQWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGKTQCARQKFS 1200
1141 FQITASGOYQWAPKLARLHYSGINAWSTKEPFSWKVDLLAPMIIHGKTQCARQKFS 1200
1201 SLVISOFTIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSGIGKNIENPPIIARYIRLHPT 1260
1201 SLVISOFTIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSGIGKNIENPPIIARYIRLHPT 1260
1261 HYSIRTLRWELMGDLNCSMPLGSKAISDAQITASSYFTNMFATWSPKARLHLQ 1320
1261 HYSIRTLRWELMGDLNCSMPLGSKAISDAQITASSYFTNMFATWSPKARLHLQ 1320
1321 RSNARPOVNNPKEMIQVDFQKTMKVTGTTQGVKSLLTSMYKKEFLISSQDGHQWTLF 1380
1321 RSNARPOVNNPKEMIQVDFQKTMKVTGTTQGVKSLLTSMYKKEFLISSQDGHQWTLF 1380
1381 FQNGKVKVQGNQDSFTPVVNSLDPLLRILHQPQSWVHQIALRMEVLGCEAODLY 1438
1381 FQNGKVKVQGNQDSFTPVVNSLDPLLRILHQPQSWVHQIALRMEVLGCEAODLY 1438

RESULT 3

US-10-225-900-1

; Sequence 1, Application US/10225900
; Publication No. US2003007752A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen

; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/10/225,900
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US/09/209,916
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-225-900-1

Query Match 100.0%; Score 7691; DB 15; Length 1438;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATRRYVILGAVELSDVWQSDLGELPVDARFPPEVPKSPFNTSVVYKTKLTFVEFTVHLN 60
DB 1 ATRRYVILGAVELSDVWQSDLGELPVDARFPPEVPKSPFNTSVVYKTKLTFVEFTVHLN 60
QY 61 IAKPRPPMGLLGPTIQAEVYTVITLKNASHPVSLHAGVSVYKASGEGAYDDQTSQ 120
DB 61 IAKPRPPMGLLGPTIQAEVYTVITLKNASHPVSLHAGVSVYKASGEGAYDDQTSQ 120
QY 121 RKEDDKVPPGSGSHYVWQVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALVCR 180
DB 121 RKEDDKVPPGSGSHYVWQVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALVCR 180
QY 181 EGS�AKEKTQTLHKFILLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
DB 181 EGS�AKEKTQTLHKFILLFAVDEGKSHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGIGCHRKSVYWHVIGMTTPVHSGIFLPGHTFLVNHROASLEISPIFLTAQTLL 300
DB 241 SLPLGIGCHRKSVYWHVIGMTTPVHSGIFLPGHTFLVNHROASLEISPIFLTAQTLL 300
QY 301 MDLQGFLLFCHLSSHQHGMAYVKVDSCEPEPOLRMKNEEAAYDDDLTSEMDVVRP 360
DB 301 MDLQGFLLFCHLSSHQHGMAYVKVDSCEPEPOLRMKNEEAAYDDDLTSEMDVVRP 360
QY 361 DDNSPSPFTQIRSVAKGHPKTVWHYIAAEEEDWDYAPLVAPDDRYSQYLNNGPQRTG 420
DB 361 DDNSPSPFTQIRSVAKGHPKTVWHYIAAEEEDWDYAPLVAPDDRYSQYLNNGPQRTG 420
QY 421 RYKVKVRWAYTDETFKTREALQHESGILGPLLYGEVGDTLIIIFKQASRPYNTYPHG 480
DB 421 RYKVKVRWAYTDETFKTREALQHESGILGPLLYGEVGDTLIIIFKQASRPYNTYPHG 480
QY 481 TDVRLYRRLPKGVKHLKDFPILRGIFPKYKMTVTVEDGPTKSDPRCLTRYSSFNME 540
DB 481 TDVRLYRRLPKGVKHLKDFPILRGIFPKYKMTVTVEDGPTKSDPRCLTRYSSFNME 540
QY 541 RDLASGLIGPLLICYKESVDQRGNQMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
DB 541 RDLASGLIGPLLICYKESVDQRGNQMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
QY 601 VQLEDPFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
DB 601 VQLEDPFQASNMHSINGVFDLSQVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKH 660
QY 661 KMVYEDTLTLFPFSGETVMSMENPGLWILCHNSDPRNGMTALLKVSSCDKNTGDIYE 720
DB 661 KMVYEDTLTLFPFSGETVMSMENPGLWILCHNSDPRNGMTALLKVSSCDKNTGDIYE 720
QY 721 DSYEDISAYLLSKNNAIEPRSFQKTRHYFIAAVERLDYGNSSPHVLRNRAQSGSVQPKVY 780
DB 721 DSYEDISAYLLSKNNAIEPRSFQKTRHYFIAAVERLDYGNSSPHVLRNRAQSGSVQPKVY 780

Db 721 DSYEDISAYLLSKNNAIEPRSFQNPVLRKHQREITRTTLQSDQBEIDYDDTISVEMKK 780
Qy 781 EDEFDIYDDEDNQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRNRAQSGSVQPKKV 840
Db 781 EDPDFIYDDEDNQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRNRAQSGSVQPKKV 840
Qy 841 FOEFTDGSFTQPLRYGBLNEHLGLLGPYIRAEVEDNIMVFRQAQSRPYFYSLLISYEE 900
Db 841 FOEFTDGSFTQPLRYGBLNEHLGLLGPYIRAEVEDNIMVFRQAQSRPYFYSLLISYEE 900
Qy 901 DQCGAGPRKMFVKPNETKTYFKVQHHMAPTDEPCKAWAFSDVLEKDVHSGLIGP 960
Db 901 DQCGAGPRKMFVKPNETKTYFKVQHHMAPTDEPCKAWAFSDVLEKDVHSGLIGP 960
Qy 961 LLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIMEDPTFK 1020
Db 961 LLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENMERNCRAPCNIMEDPTFK 1020
Qy 1021 NYRPHALNGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIHSGHVFTVRKERYMA 1080
Db 1021 NYRPHALNGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIHSGHVFTVRKERYMA 1080
Qy 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGNASGHIRD 1140
Db 1081 LYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCQTPLGNASGHIRD 1140
Qy 1141 FQITASQYQWAPKLARLYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTGARQKFS 1200
Db 1141 FQITASQYQWAPKLARLYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTGARQKFS 1200
Qy 1201 SLVISOPIINSLDGKWKQYRGNSGTGLMVFVGNVDSSGINKINFPPIIARIHPT 1260
Db 1201 SLVISOPIINSLDGKWKQYRGNSGTGLMVFVGNVDSSGINKINFPPIIARIHPT 1260
Qy 1261 HYSIRSTLRMELMGCDLNSCMLPGMESKAISSAQITASSYFTNMFTATWSPSKARHLQ 1320
Db 1261 HYSIRSTLRMELMGCDLNSCMLPGMESKAISSAQITASSYFTNMFTATWSPSKARHLQ 1320
Qy 1321 RSNARPOVNNKLEWLDVQDKTMKVTGVTQGVKSLLTSMYKFEFLISSQDGHQWTLF 1380
Db 1321 RSNARPOVNNKLEWLDVQDKTMKVTGVTQGVKSLLTSMYKFEFLISSQDGHQWTLF 1380
Qy 1381 FQNGKVKVFCQNDSPFPVNSLDPLLRILYRHPQSWHQAIALRMEVLGCEAQDLY 1438
Db 1381 FQNGKVKVFCQNDSPFPVNSLDPLLRILYRHPQSWHQAIALRMEVLGCEAQDLY 1438

RESULT 4

US-10-095-718-2
; Sequence 2, Application US/10095718
; Publication No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCES: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII

; FEATURE:

; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2

Query Match 99.8%; Score 7674; DB 14; Length 1471;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 14; Gaps 1;

Qy 1 ATRRYVLGAVELSWDMQSDLGELPVDASPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
Db 20 ATRRYVLGAVELSWDMQSDLGELPVDASPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 79
Qy 61 IAKPRPPMGLGPTTQAEVYDTVTTLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 120
Db 80 IAKPRPPMGLGPTTQAEVYDTVTTLKNMASHPVSLHAGVSVYKASEGAEYDDQTSQ 139
Qy 121 RKEDDKVPFGASHVTVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
Db 140 RKEDDKVPFGASHVTVMQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 199
Qy 181 EGS�AKBKQTTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 240
Db 200 EGS�AKBKQTTLHKFTLLFAVDEGKSWHSETKNSLMQDRDAASARAMPKMTVNGYVNR 259
Qy 241 SLPLIGCHRKSVYHVIHGMGTTPVHSIFLEGHTPLVNRHQASLEISPIITFLTAQTLL 300
Db 260 SLPLIGCHRKSVYHVIHGMGTTPVHSIFLEGHTPLVNRHQASLEISPIITFLTAQTLL 319
Qy 301 MDLGOPLFLFCHTSSHQHDMBAVVKVDSCEPQPQRMKNBERAEDYDDDLTDTSEMDVVRP 360
Db 320 MDLGOPLFLFCHTSSHQHDMBAVVKVDSCEPQPQRMKNBERAEDYDDDLTDTSEMDVVRP 379
Qy 361 DDNSPSFTQIRSVAKKHPKTVVHYIAAEEEDMDYAPLVLAPEDDRSYKSQYLNNQPRIG 420
Db 380 DDNSPSFTQIRSVAKKHPKTVVHYIAAEEEDMDYAPLVLAPEDDRSYKSQYLNNQPRIG 439
Qy 421 RYKVKRWAYTDEFTKREATQESGILGPLLYGEVGDTLIIIFKNQASRPNTYPHGI 480
Db 440 RYKVKRWAYTDEFTKREATQESGILGPLLYGEVGDTLIIIFKNQASRPNTYPHGI 499
Qy 481 TDVRPLYSRRLPKGVKHLKDPFIPILGETFKYKWTVTVEGPTKSPRCITRYYSFVWME 540
Db 500 TDVRPLYSRRLPKGVKHLKDPFIPILGETFKYKWTVTVEGPTKSPRCITRYYSFVWME 559
Qy 541 RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLIFSVDENRSLYLTENIQRFLEPNAG 600
Db 560 RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLIFSVDENRSLYLTENIQRFLEPNAG 619
Qy 601 VQLEDPEFQASNMHSINGYVDSLOLSVCLHEVAYWYILSIGATDFLSVFFSGYTKH 660
Db 620 VQLEDPEFQASNMHSINGYVDSLOLSVCLHEVAYWYILSIGATDFLSVFFSGYTKH 679
Qy 661 KMVYEDTLTLFPFSGETVFMGMENFGLMILGCHNSDFNRNMGMTALLKYSSCDKNTGDYVE 720
Db 680 KMVYEDTLTLFPFSGETVFMGMENFGLMILGCHNSDFNRNMGMTALLKYSSCDKNTGDYVE 739
Qy 721 DSYEDISAYLLSKNNAIEPRSFQNPVLRKHQREITRTTLQSDQBEIDYDDTISVEMKK 766
Db 740 DSYEDISAYLLSKNNAIEPRSFQNPVLRKHQREITRTTLQSDQBEIDYDDTISVEMKK 799
Qy 767 EIDYDDTISVEMKKEDFDIYDEDNQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRN 826
Db 800 EIDYDDTISVEMKKEDFDIYDEDNQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRN 859
Qy 827 RAQSGSVQPKVQFQFTDGSFTQPLRYGBLNEHLGLLGPYIRAEVEDNIMVFRQAQ 886
Db 860 RAQSGSVQPKVQFQFTDGSFTQPLRYGBLNEHLGLLGPYIRAEVEDNIMVFRQAQ 919
Qy 887 RPYSFYSSLSIYEDDQAGAPRKNFVKPNETKTYFKVQHHMAPTDEPCKAWAFSD 946
Db 920 RPYSFYSSLSIYEDDQAGAPRKNFVKPNETKTYFKVQHHMAPTDEPCKAWAFSD 979
Qy 947 VDLEKDVHSGLIGLPLLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENMERNCR 1006

Db 980 VDLEKDVHSLGILLPVLVCHTNTLPAHGRQVTVQEPALFFTFIDETKSWYFTENMERNCK 1039
QY 1007 APCNIQWEDPTFKENTRFHAINGYINDTLPLGLVMAQDQIRWYLLSNGSNEINHSIHFSG 1066
Db 1040 APCNIQWEDPTFKENTRFHAINGYINDTLPLGLVMAQDQIRWYLLSNGSNEINHSIHFSG 1099
QY 1067 HFTVTRKCEYKMAIYNLYPGVPFETVEMLPFSKAGIWRVECLIGEHLAGMSTLFLVYSNK 1126
Db 1100 HFTVTRKCEYKMAIYNLYPGVPFETVEMLPFSKAGIWRVECLIGEHLAGMSTLFLVYSNK 1159
QY 1127 CQTPPLWASHIRDFQITASQYQWAPKARLHYSGSINASTKEPFPFWKVDLLAPMI 1186
Db 1160 CQTPPLWASHIRDFQITASQYQWAPKARLHYSGSINASTKEPFPFWKVDLLAPMI 1219
QY 1187 IHGIKTQARQKPFSSLYISQFIIMYSIDGKQWYRGNSGTGLMVFFGNVDSGIGKHNP 1246
Db 1220 IHGIKTQARQKPFSSLYISQFIIMYSIDGKQWYRGNSGTGLMVFFGNVDSGIGKHNP 1279
QY 1247 NPPIIARYIRLHPHTHYSIRSLRMELMGCDLNSCSMPLQWESKAIISDAQITASSYFTNMF 1306
Db 1280 NPPIIARYIRLHPHTHYSIRSLRMELMGCDLNSCSMPLQWESKAIISDAQITASSYFTNMF 1339
QY 1307 ATWSPSKARLHLOGRSNARFPQNNPKWLQVDFQKTMKVTGTTQGVKSLTSMYVKEF 1366
Db 1340 ATWSPSKARLHLOGRSNARFPQNNPKWLQVDFQKTMKVTGTTQGVKSLTSMYVKEF 1399
QY 1367 LISSSQDGHQWTLFFQNGKVKVYFQGNQDSFTPVNSLDPLLTTRYLRHQPQSWHQAIALR 1426
Db 1400 LISSSQDGHQWTLFFQNGKVKVYFQGNQDSFTPVNSLDPLLTTRYLRHQPQSWHQAIALR 1459
QY 1427 MEVLGCEAODLY 1438
Db 1460 MEVLGCEAODLY 1471

RESULT 5

US-09-957-641-2

; Sequence 2, Application US/09957641

; Publication No. US20020182670A1

; GENERAL INFORMATION:

; APPLICANT: Emory University

; TITLE OF INVENTION: MODIFIED FACTOR VIII

; FILE REFERENCE: 75-00

; CURRENT APPLICATION NUMBER: US/09/957,641

; CURRENT FILING DATE: 2001-09-16

; PRIOR APPLICATION NUMBER: US 60/234047

; PRIOR FILING DATE: 2000-09-19

; PRIOR APPLICATION NUMBER: US 60/236460

; PRIOR FILING DATE: 2000-09-29

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-957-641-2

Query Match

Best Local Similarity 94.1%; Score 7234; DB 10; Length 2332;

Matches 1438; Conservative 61.7%; Pred. No. 0;

Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRRYVLGAVELSWDYNQSDIGELPVDARFPDPRPKSPFPNFSVYKXKTLFVFEFTVHLFN 60

Db 1 ATRRYVLGAVELSWDYNQSDIGELPVDARFPDPRPKSPFPNFSVYKXKTLFVFEFTVHLFN 60

QY 61 IAKRPPPNWGLLGPITQAEVVDTVVITLKNWASHVPVSLHAYGVSWKASEGAEVDDQTSQ 120

Db 61 IAKRPPPNWGLLGPITQAEVVDTVVITLKNWASHVPVSLHAYGVSWKASEGAEVDDQTSQ 120

QY 121 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180

Db 121 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180

QY 181 EGSIAKCKTQTLHKETLLPAPVDECKSWHSETKNSIMQDRDAASARAWPKMTYVYNR 240
Db 181 EGSIAKCKTQTLHKETLLPAPVDECKSWHSETKNSIMQDRDAASARAWPKMTYVYNR 240
QY 241 SLPLGLIGCHRSVYVHVHVGMTTPEVHSIFLEHHTFLVNRHQASLEISPIITFLTAOTLL 300
Db 241 SLPLGLIGCHRSVYVHVHVGMTTPEVHSIFLEHHTFLVNRHQASLEISPIITFLTAOTLL 300
QY 301 MDLQGLFLPCHTSSHOHDGMEAYVVDSCPEBPQLRMKNNEBAYDDDLTDSMDVVVF 360
Db 301 MDLQGLFLPCHTSSHOHDGMEAYVVDSCPEBPQLRMKNNEBAYDDDLTDSMDVVVF 360
QY 361 DDNSPSTQIERSVAKKPKTWVHYTAAEEDWDYAPLVLADDDRSYKQYLNGPQRTG 420
Db 361 DDNSPSTQIERSVAKKPKTWVHYTAAEEDWDYAPLVLADDDRSYKQYLNGPQRTG 420
QY 421 RYKVKVRFWAYTDETFKTREAIQHESGILGLPILLYGEVDTLLIIFKNOASRPYNYPHGI 480
Db 421 RYKVKVRFWAYTDETFKTREAIQHESGILGLPILLYGEVDTLLIIFKNOASRPYNYPHGI 480
QY 481 TDVRLYKRRPLPKGVKHLKDPFLLPGEIFKVKWTVVDEGPTKSDPCLTRYSSFFVME 540
Db 481 TDVRLYKRRPLPKGVKHLKDPFLLPGEIFKVKWTVVDEGPTKSDPCLTRYSSFFVME 540
QY 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKENVILFSVFDENRSHVLTENTQRFNPAG 600
Db 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKENVILFSVFDENRSHVLTENTQRFNPAG 600
QY 601 VOLEDPEFQASINHSINGYVFDLSIQLSVCLHEVAYWYILSIGACTDPLSVFFSYTFKH 660
Db 601 VOLEDPEFQASINHSINGYVFDLSIQLSVCLHEVAYWYILSIGACTDPLSVFFSYTFKH 660
QY 661 KMYVEDTLTLFPFSGETVFMENPGLMWILGCHNSDFRNKGWMTALLKVVSCDXTGDIYE 720
Db 661 KMYVEDTLTLFPFSGETVFMENPGLMWILGCHNSDFRNKGWMTALLKVVSCDXTGDIYE 720
QY 721 DSYEDISAYLLSKNAIEPRSF----- 742
Db 721 DSYEDISAYLLSKNAIEPRSFQNSRHPSTROKQFNATTIPENDIEKTDPAFHTMP 780
QY 743 ----- 742
Db 781 KIQNVSSDLMLARQSPTHGLSLSDLOEAKYETFPDDPSPGAIDSNNSLSEMTFRPQ 840
QY 743 ----- 742
Db 841 LHSQDMVYTPESGLQLRNEKLGTTAATLKLDFKVSSTNNLISTIPSDNLAAGTDN 900
QY 743 ----- 742
Db 901 TSSLGPPSPMPHYDSQDLDTTLFGKSSPLTESGGPLSEENNDKLLSGLMNSQESSW 960
QY 743 ----- 742
Db 961 GKNVSTESGRLFKGKRAHGPAALLTKDNALFKVISLTKTKTNSNATNPKTHIDGPSL 1020
QY 743 ----- 742
Db 1021 LIENSPVWQNILESDFEKKVTPLIHDMRLMDKNATALRLNHNMSNKTSSKNMEMVQOK 1080
QY 743 ----- 742
Db 1081 KEGPIPPDAQNPDSFFKMLFLPESARWIQRTGKNSLNSGQSPSPKQVLSLGEKSVEG 1140
QY 743 ----- 742
Db 1141 QNFLSERKNVVGKGEFTKDVGLKEMVFPSSRNLFLTNLDLHNHNTHNQEKKIQEBIEK 1200
QY 743 ----- 742
Db 1201 KETLIQENVVLPOIHTVTGTGKNFMKNLFLLLSTRQNVESGYEGAYAPVLQDPRSLDSTNR 1260


```
QY 743 ----- 742
Db 1261 TKKHTAHFSKGBEENLEGLGNQTKQIVKIACTTRISPTSQNFVQTSKRALKQFRL 1320
QY 743 ----- 742
Db 1321 PLEETELEKRIIYDDTSTQSKNMKHLTSTLTQIDYNEKEKAITQSPSLDCLTRSHI 1380
QY 743 ----- 742
Db 1381 POANRSPLEIAKYSPFSIRPIYLRVLFDQNSSHLPAASRYKKGDSGVQESSHFLQAKK 1440
QY 743 ----- 742
Db 1441 NNLSIALILEMTGDQREVSLGTSATNSVYKXKVENTVLPKPDLPKTSKGVELLPKVHI 1500
QY 743 ----- 742
Db 1501 YQKDLFPPTETSNPGCHLDLVEGSLAQTEGAIKWNEANRPGKVPPLRVATRESSAKTPSK 1560
QY 743 ----- 742
Db 1561 LLDPLANDNHGTQIPKEBWSQESPEKTAFAKKDITLSLNACSNHAIANAINEQONKP 1620
QY 743 ----- 786
Db 1621 EIEVTWAKQGRTERLCSQPPVLRKHQREIRITTLQSDQEEIDYDDTISVENMKGBDFDIY 1680
QY 787 DEBENQSPRSFQKTRHYFAIAVERLWDYGMSSSPHVLNRRAQSGSVPOFKKVVQFEFTD 846
Db 1681 DEBENQSPRSFQKTRHYFAIAVERLWDYGMSSSPHVLNRRAQSGSVPOFKKVVQFEFTD 1740
QY 847 GSFTQPLVGELEHGLGPIYARVARDYINIMVTFRNQASRPSPYSSLSIYEBORQQA 906
Db 1741 GSFTQPLVGELEHGLGPIYARVARDYINIMVTFRNQASRPSPYSSLSIYEBORQQA 1800
QY 907 EPRKNFVKNETKTYFWKQHMATPKDBFDCKAWAYFSDVLEKDVHSGGLGPIVLVCHT 966
Db 1801 EPRKNFVKNETKTYFWKQHMATPKDBFDCKAWAYFSDVLEKDVHSGGLGPIVLVCHT 1860
QY 967 NTLNPAHQVQVTEFALFTTIDETKSWYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1026
Db 1861 NTLNPAHQVQVTEFALFTTIDETKSWYFTENMERNCRAPCNQIMEDPTFKENYRPHA 1920
QY 1027 INGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTRKKBEYKQALYNLYP 1086
Db 1921 INGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTRKKBEYKQALYNLYP 1980
QY 1087 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDFQITAS 1146
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMASGHIRDFQITAS 2040
QY 1147 GYGOWAPKLABLHYSGSINASTKEPFSWIKVOLLAPMIINGIKTQARQKFSLLYISQ 1206
Db 2041 GYGOWAPKLABLHYSGSINASTKEPFSWIKVOLLAPMIINGIKTQARQKFSLLYISQ 2100
QY 1207 FIIMYSLDGKKQVTRGNSGTILWVFFGNVDSGIGKNIENPPIIARYIRLPHTHYSIRS 1266
Db 2101 FIIMYSLDGKKQVTRGNSGTILWVFFGNVDSGIGKNIENPPIIARYIRLPHTHYSIRS 2160
QY 1267 TURBELMGCDLNSCNPGLMESKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 1326
Db 2161 TURBELMGCDLNSCNPGLMESKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 2220
QY 1327 POWNPKWLQVDFOKTMKVGTVTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 1386
Db 2221 POWNPKWLQVDFOKTMKVGTVTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
QY 1387 KVFQGNQDSFTFVNASLDPPLLTRYLIRHPQSWVHQIALRMEVLGCEAQDLY 1438
Db 2281 KVFQGNQDSFTFVNASLDPPLLTRYLIRHPQSWVHQIALRMEVLGCEAQDLY 2332
```

RESULT 6

```
US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131.510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-510A-2
```

```
Query Match 94.1%; Score 7234; DB 12; Length 2332;
Best Local Similarity 61.7%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 894; Gaps 1;

QY 1 ATRRYLCAVRLSDWDMQSDGLGELPVDARFPVPKSPPTSVVYKTLVFEFVHLFN 60
Db 1 ATRRYLCAVRLSDWDMQSDGLGELPVDARFPVPKSPPTSVVYKTLVFEFVHLFN 60
QY 61 IAKPRPPWGLLGPTIQAEVVDYTVITLKNMASHPVSLHVGVSYKASEGAEVDDQTSQ 120
Db 61 IAKPRPPWGLLGPTIQAEVVDYTVITLKNMASHPVSLHVGVSYKASEGAEVDDQTSQ 120
QY 121 REKEDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 180
Db 121 REKEDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 180
QY 181 EGSLAKEKTQTLHKFILLFAVDEGKSWHSETKSLMDRDAASARAWPKMHTVNGYVNR 240
Db 181 EGSLAKEKTQTLHKFILLFAVDEGKSWHSETKSLMDRDAASARAWPKMHTVNGYVNR 240
QY 241 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEHGTFLVRNHRQASLEISPTITLTAOTLL 300
Db 241 SLPLGLICHRKSVVWHVIGMTTPEVHSIFLEHGTFLVRNHRQASLEISPTITLTAOTLL 300
QY 301 MDLQGFLLFCHISSHQDGMAYVKVDSCEPQPMKNNBEAEYDDDLTDSMDVVVRP 360
Db 301 MDLQGFLLFCHISSHQDGMAYVKVDSCEPQPMKNNBEAEYDDDLTDSMDVVVRP 360
QY 361 DDDNSPSPFIQIRSVAKKHPKTTWHVIAAREEDWDYAPLVLAPDORSYKSQLNNGPORIG 420
Db 361 DDDNSPSPFIQIRSVAKKHPKTTWHVIAAREEDWDYAPLVLAPDORSYKSQLNNGPORIG 420
QY 421 RYKVKRPMAYTDTFTKTRAIQHSGLIGPLLYGEGVDTLIIIFKKNQASRPYNIYPHGI 480
Db 421 RYKVKRPMAYTDTFTKTRAIQHSGLIGPLLYGEGVDTLIIIFKKNQASRPYNIYPHGI 480
QY 481 TDVREPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPCLTRYYSFVWME 540
Db 481 TDVREPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPCLTRYYSFVWME 540
QY 541 RDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG 600
```

Db 541 RDLASGLIGLPLLCYKESVDQGNQJMSDKRNVLFSVPDENRSWLTENIQEFLENPAG 600
QY 601 VOLEDEFFOASNTMHSINGVVDLSQLSVCLHEVAYWYILISCAQTDPLSVFSGTTFKH 660
Db 601 VOLEDEFFOASNTMHSINGVVDLSQLSVCLHEVAYWYILISCAQTDPLSVFSGTTFKH 660
QY 661 KMVYEDTLTLFPFSGETVFMMSMENPGLWILGCHNSDFRNGMTALLKVSJCDKNTGDIYE 720
Db 661 KMVYEDTLTLFPFSGETVFMMSMENPGLWILGCHNSDFRNGMTALLKVSJCDKNTGDIYE 720
QY 721 DSYEDISAYLLSKNAIEPRSP 742
Db 721 DSYEDISAYLLSKNAIEPRSP 742
QY 743 742
Db 781 KIQNVSSDLLMLLRQSPPHGLSLSDLQBAKYETSDDDSPGAIIDSNNSLSWTHFRQ 840
QY 743 742
Db 841 LHSGDMVFTPESGLQLRLNEKLGTTAATLKLDFKVSSTSNLIJSTIPSDNLAAGTDN 900
QY 743 742
Db 901 TSSLGPPSPVHVDSQDITLFGKKSPLTESGGPLSLEENNDKLSGLNSQESSW 960
QY 743 742
Db 961 GKNVSTESGRLPFGKRAHGPALLTKDNALFKVISILKTKNTSNNSATNRKTHIDGPSL 1020
QY 743 742
Db 1021 LIENSPSPVQNILESDETFKKTPLIHDRMLDKNATLRLNHSNKTSSKNMENVQOK 1080
QY 743 742
Db 1081 KEGPIPPDQNDPMSFFKMLFLPESARNIQRTHGRKNSLNSGGQSPKQVLVSLGPEKSVEG 1140
QY 743 742
Db 1141 QNFLEKKNVVGKGBFTXVGLKEMVFPSSRNFLTNDLNLHENHNTNQEKKIQBIETK 1200
QY 743 742
Db 1201 KETLIOENVVLPOIHTVTGTGNFMKNLFLSTRQNVGSEYAGAPVLQDFRSLNDSTR 1260
QY 743 742
Db 1261 TKKHTAHPKSGEENLEGLGNQTKQIVBKACTTRISPTSQONFVTQORSKRALKQFRL 1320
QY 743 742
Db 1321 PLEETELEXRIIIVDDTSTQSKMKHLPSTLTQIDYNEKEGATQSPSLDCLTRSHSI 1380
QY 743 742
Db 1381 POANRSPLEIAKVSSFPRIPIYLRVLFPQDNSHLPAAZYRKDQSGVQESSHFLQGAOK 1440
QY 743 742
Db 1441 NNLSLAILTEMTQGRVSGSLGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVIH 1500
QY 743 742
Db 1501 YOKDLFPETSTNGSPHLDLVEGLIQTEGAIKWNEARNPKGVPLRVATESSAKTPSK 1560
QY 743 742
Db 1561 LLDPLAWNHYGTQIPKEWKSQESKPEKATFKKDDTILSNACSNHAIARINEGQNK 1620
QY 743 742

Db 1621 ELEVWAKQGRTERLCSQNPPLVKQRHREITRTTLQSDQBEIDYDDTISVEMKCEDFDIY 1680
QY 787 DDEBNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAGSGSVPOFKKVVFOEFTD 846
Db 1681 DDEBNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAGSGSVPOFKKVVFOEFTD 1740
QY 847 GSFTQPLVYRGELEHNLGLGPYIRAEVEDNIWVTPRNOASRPVSPVSSLSISVEEDORQA 906
Db 1741 GSFTQPLVYRGELEHNLGLGPYIRAEVEDNIWVTPRNOASRPVSPVSSLSISVEEDORQA 1800
QY 907 EPRKNFVQNETKTYFMKVOHMAPTKDFDCKAMAYFSDVLEKDVHSGSLGLPLLVCHT 966
Db 1801 EPRKNFVQNETKTYFMKVOHMAPTKDFDCKAMAYFSDVLEKDVHSGSLGLPLLVCHT 1860
QY 967 NTLNPAHQROVTVQBEALFFETIDETKSHYFTENNERCBAQNTQMBDPTKENYRFA 1026
Db 1861 NTLNPAHQROVTVQBEALFFETIDETKSHYFTENNERCBAQNTQMBDPTKENYRFA 1920
QY 1027 INGYIMDTLPLGLVMAQDQIRIWMYLLSMGNSNENIHSIFSGHVFTVRKKEEYKMALYNLYP 1086
Db 1921 INGYIMDTLPLGLVMAQDQIRIWMYLLSMGNSNENIHSIFSGHVFTVRKKEEYKMALYNLYP 1980
QY 1087 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKQCTPLGMAHSHIRDFQITAS 1146
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKQCTPLGMAHSHIRDFQITAS 2040
QY 1147 GYGOWAPKARLHYSGSINAWSTKEPFSWIKVLLAPMIHGKITQAGROKESLYISO 1206
Db 2041 GYGOWAPKARLHYSGSINAWSTKEPFSWIKVLLAPMIHGKITQAGROKESLYISO 2100
QY 1207 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSGKIKHIFNPPIIARILRHPTHYSIRS 1266
Db 2101 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSGKIKHIFNPPIIARILRHPTHYSIRS 2160
QY 1267 TURBELMGCDLNSCMLPGLMESKALSDAQITASSYFTNMFTWSPSKARLHLOGRSNWR 1326
Db 2161 TURBELMGCDLNSCMLPGLMESKALSDAQITASSYFTNMFTWSPSKARLHLOGRSNWR 2220
QY 1327 PQVNNPKELQVDFQTKMKVTGVTQGVKSLTSMYKEFLISSQDGHOWTLFPQNGKV 1386
Db 2221 PQVNNPKELQVDFQTKMKVTGVTQGVKSLTSMYKEFLISSQDGHOWTLFPQNGKV 2280
QY 1387 KVFQGNQDSFTFVNVNSLDPPLITRYLRIRHPSQSWHQAIALRMEVLGCEAQDLY 1438
Db 2281 KVFQGNQDSFTFVNVNSLDPPLITRYLRIRHPSQSWHQAIALRMEVLGCEAQDLY 2332

RESULT 7

US-10-187-319-2
; Sequence 2, Application US/10187319
; Publication No. US20030068785A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,656

787 DEDNQSPRSQKXTRHYFIAAVERLMDYGMSSSPHVLNRRAQSGVSPQKVVQFBFTD 846
 1681 DEDNQSPRSQKXTRHYFIAAVERLMDYGMSSSPHVLNRRAQSGVSPQKVVQFBFTD 1740
 847 GSPTQPLXRGELNEHLGLGPIYIRAEVEDNIMWTFRNQASRPYSFYSSLSYEBDQDGA 906
 1741 GSFTQPLXRGELNEHLGLGPIYIRAEVEDNIMWTFRNQASRPYSFYSSLSYEBDQDGA 1800
 907 EPRKNFVKPNETKTYFNKVOHMAPTKDEPCKAWAYFSDVLEKDVHSGLIGLPLVCHT 966
 1801 EPRKNFVKPNETKTYFNKVOHMAPTKDEPCKAWAYFSDVLEKDVHSGLIGLPLVCHT 1860
 967 NTLNPAHGRQVTVQEFALFTTFDETKSWYFTENMERNCRAPCIOMEDPTFKENYRFA 1026
 1861 NTLNPAHGRQVTVQEFALFTTFDETKSWYFTENMERNCRAPCIOMEDPTFKENYRFA 1920
 1027 INGVIYMDTLPLCLVMAQDRIWYLLSMGSENTHSIFHSGHVFYRKKKEYKVALYNLYP 1086
 1921 INGVIYMDTLPLCLVMAQDRIWYLLSMGSENTHSIFHSGHVFYRKKKEYKVALYNLYP 1980
 1087 GVFTETVEMLPKSKAGIMRVECLIGELHAGMSTLFLVYSNKCQTPGLMASGHIRDFOITAS 1146
 1981 GVFTETVEMLPKSKAGIMRVECLIGELHAGMSTLFLVYSNKCQTPGLMASGHIRDFOITAS 2040
 1147 GOYGOWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKPSLSYISQ 1206
 2041 GOYGOWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQARQKPSLSYISQ 2100
 1207 FIWYSLDGKKWQTYRGNSTGLMVFPGNDVSSGKHNINFPPIIARYIRLHPHYSIRS 1266
 2101 FIWYSLDGKKWQTYRGNSTGLMVFPGNDVSSGKHNINFPPIIARYIRLHPHYSIRS 2160
 1267 TLRMELMGCDLNSCPLMGESKASIDAQITASSYFTNMFATWSPSKARLHLQGRNNAWR 1326
 2161 TLRMELMGCDLNSCPLMGESKASIDAQITASSYFTNMFATWSPSKARLHLQGRNNAWR 2220
 1327 PQVNNPKWLOVDFOKTMKVTGTTQGVKSLLTSMYKPELISSSQDGHQWTLFFQNGKV 1386
 2221 PQVNNPKWLOVDFOKTMKVTGTTQGVKSLLTSMYKPELISSSQDGHQWTLFFQNGKV 2280
 1387 KVFQGNQDSFTPVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 1438
 2281 KVFQGNQDSFTPVNSLDPLLRILRIHPQSWHQAIALRMEVLGCEAODLY 2332

RESULT 8

US-10-133-907-4
 ; Sequence 4, Application US/10133907
 ; Publication No. US20030195223A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chien, Kenneth R
 ; APPLICANT: Hoshijima, Masahiko
 ; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII
 ; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII
 ; FILE REFERENCE: 6627-Pall170
 ; CURRENT APPLICATION NUMBER: US/10/133,907
 ; PRIOR FILING DATE: 2002-04-25
 ; PRIOR APPLICATION NUMBER: 60/286,314
 ; PRIOR FILING DATE: 2001-04-25
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 2351
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-133-907-4

Query Match 94.0%; Score 7227; DB 12; Length 2351;
 Best Local Similarity 61.6%; Pred. No. 0;
 Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;
 1 ATRRYYLGAVELSMDYQSDLGELPVDARFPPRPVKSPFNTSVVYKKTFLVEFTVHLFN 60

Db 20 ATRRYYLGAVELSMDYQSDLGELPVDARFPPRPVKSPFNTSVVYKKTFLVEFTVHLFN 79
 QY 61 IAKPRPPWMLLQPTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
 Db 80 IAKPRPPWMLLQPTQAEVYDVTWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 139
 QY 121 REKEDDKVPFGGSHYTVWVLYKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 Db 140 REKEDDKVPFGGSHYTVWVLYKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
 QY 181 EGSIAKEKQTTLHKFTILLFAVDFEGSKSWHSETKNSLMQDRDAASARAWPKGHTVNGVNR 240
 Db 200 EGSIAKEKQTTLHKFTILLFAVDFEGSKSWHSETKNSLMQDRDAASARAWPKGHTVNGVNR 259
 QY 241 SLPLGLGCHRSKYVHWVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300
 Db 260 SLPLGLGCHRSKYVHWVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 319
 QY 301 MDLQGFLLFCHISSHODGWEAVKVDSCPEEPQLRMKNEAEYDDDLTDSMDVVRP 360
 Db 320 MDLQGFLLFCHISSHODGWEAVKVDSCPEEPQLRMKNEAEYDDDLTDSMDVVRP 379
 QY 361 DDNSPSFIQIRSAVAKHPKTKWVHYTAAEEEDMDYAPLVLAPDDRYSKYQLNNGPQIRG 420
 Db 380 DDNSPSFIQIRSAVAKHPKTKWVHYTAAEEEDMDYAPLVLAPDDRYSKYQLNNGPQIRG 439
 QY 421 RYKVKYRWAYTDETPKTRATOHESGILGPLLYGEGVDTLLIIFKNQASRPNIYPHGI 480
 Db 440 RYKVKYRWAYTDETPKTRATOHESGILGPLLYGEGVDTLLIIFKNQASRPNIYPHGI 499
 QY 481 TDVRLYSLRPLKGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYVSSVFNME 540
 Db 500 TDVRLYSLRPLKGVKHLKDFPILPGEIFKYKWTVTVEGPTKSDPRCLTRYVSSVFNME 559
 QY 541 RLASGLIGPLIICYKESVDQRGNQIMSDKRNVLIFSVDENRSWYLTENIQRFLENPAG 600
 Db 560 RLASGLIGPLIICYKESVDQRGNQIMSDKRNVLIFSVDENRSWYLTENIQRFLENPAG 619
 QY 601 VOLEPPEQASNIMHSINGYVDSLSQVCLHEVAYWYILSIGAQTDFLSVFPFGYTFKH 660
 Db 620 VOLEPPEQASNIMHSINGYVDSLSQVCLHEVAYWYILSIGAQTDFLSVFPFGYTFKH 679
 QY 661 KXVYEDTLTLFPFSGETVFMSENGLMILGCHNSDFRNRGMTALKVSSCDKNIGDYIE 720
 Db 680 KXVYEDTLTLFPFSGETVFMSENGLMILGCHNSDFRNRGMTALKVSSCDKNIGDYIE 739
 QY 721 DSYEDISAYLLSKNNAIEPRSP- 742
 Db 740 DSYEDISAYLLSKNNAIEPRSP- 799
 QY 743 742
 Db 800 KIQNVSSDMLLRQSPTPHQLSUSDLQAEKYEFTSDPSPGAIDSNNSISEMTHFRPQ 859
 QY 743 742
 Db 860 LHHSGDMVTFPESGLQLRLNEKLGTTAATELKCLDFKVSSISNNLISIPSDNLAAGTDN 919
 QY 743 742
 Db 920 TSSLGPPSPMVHYDSQDITLFGKKSPLTBSGGLSLEENNDKLEESGLMNSQESSW 979
 QY 743 742
 Db 980 GKNVSTESGRLFKGKRAHGAPALLTKONALPKVISLTKNTKNSNATNRKTHIDGPSL 1039
 QY 743 742
 Db 1040 LIENSPPVWQNLISDTEFKVTPILLHDBMLMDXNATRLNHNHNSKNTTSSKNEMWQOK 1099
 QY 743 742
 Db 1100 KEGPIPPDAQNPDMSPFFKMLFPPESARWIORTHGKNSLNSQGPSPKQLVSLGPEKSV 1159

```
QY 743 ----- 742
Db 1160 QNPLSEKNVVGKGEFTKDVGLKEMVFPSSRNFLTNLNLHNNTHNQEKIQBIEK 1219
QY 743 ----- 742
Db 1220 KETLIQENVVLPOIHTVTGKNFMKNLFLSTRQNVGSDGAYAPVLQDPRSLNDSTNR 1279
QY 743 ----- 742
Db 1280 TKKHTAHFKKGBENLEGLNQTQKIVKACTTRISPNTSQNFVTVQSRKALKQPRL 1339
QY 743 ----- 742
Db 1340 PLBETELKRIIVDDTQWSSKNMKHLTPSTLTQIDYNEKEKGAITQPSLSDCLTRSHSI 1399
QY 743 ----- 742
Db 1400 PQANRSLPIAKVSSFPPIRPIYLTRVLFDQNSHLPAAVSRKDSGVQESSHPLQAKK 1459
QY 743 ----- 742
Db 1460 NNLSLAILTLEMTQDQREVSLGTSATNSVTYKKVENTVLKPDLPKTSKGVKLLPKVHI 1519
QY 743 ----- 742
Db 1520 YQKDLFTTETNSGPHLDLVEGSLQGTGEGAIKWEANRPGKVPFLRVATESAKTPSK 1579
QY 743 ----- 742
Db 1580 LLDPLAWNHYGTQIPKEBWKSEKSPKTAFFKKDKTILSLNACSNHAIANEQNK 1639
QY 743 ----- 786
Db 1640 EIEVTWAKQGRTERLCSQNPFLKQHREITRTLTQSDQBEIDYDDTISYEMKEDFDIY 1699
QY 787 DEBENQSPRSFQKTRHYFAAVERLWDYQSSPHVLRNRAQSGVPOPKVVOBEFTD 846
Db 1700 DEBENQSPRSFQKTRHYFAAVERLWDYQSSPHVLRNRAQSGVPOPKVVOBEFTD 1759
QY 847 GSTQPLYRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYBEDORQA 906
Db 1760 GSTQPLYRGELNEHLLGPGYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYBEDORQA 1819
QY 907 EPRKFNVPNETKTYFKVQOHMAPTKDEPDKAWAYFSDVDLEKDVHSGLIQPLVCHT 966
Db 1820 EPRKFNVPNETKTYFKVQOHMAPTKDEPDKAWAYFSDVDLEKDVHSGLIQPLVCHT 1879
QY 967 NTLNPAHGRQVTVQEPALFPTTIDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA 1026
Db 1880 NTLNPAHGRQVTVQEPALFPTTIDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRFA 1939
QY 1027 INGYINDTLRGUWAOQRIWYLLSGNSNENHSHFSGHVTFVTKKSEYKQALVNLYP 1086
Db 1940 INGYINDTLRGUWAOQRIWYLLSGNSNENHSHFSGHVTFVTKKSEYKQALVNLYP 1999
QY 1087 GVPETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLQWASHLRDFOITAS 1146
Db 2000 GVPETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLQWASHLRDFOITAS 2059
QY 1147 GOYGWAPKLARLHYSGSINAMSTKEPFSWKVDLLAPMI IHGKITQGARQKFSSLIYISQ 1206
Db 2060 GOYGWAPKLARLHYSGSINAMSTKEPFSWKVDLLAPMI IHGKITQGARQKFSSLIYISQ 2119
QY 1207 FIIMYSLDGKKQWTVGNGTGLMWFGNVDSGIGKNI FNPPIIARYIRLHPHTYSIRS 1266
Db 2120 FIIMYSLDGKKQWTVGNGTGLMWFGNVDSGIGKNI FNPPIIARYIRLHPHTYSIRS 2179
QY 1267 TLRMELMGCDLNSCSNPLGWESKAI SDAQITASSYFTNMFWATWSPSKARLHQGRSNAR 1326
Db 2180 TLRMELMGCDLNSCSNPLGWESKAI SDAQITASSYFTNMFWATWSPSKARLHQGRSNAR 2239
```

```
QY 1327 PQVNNPREMLQVDFQTKMKVTGTTQGVKSLTSMYVKEFLISSSQDGHQWTLFFQNGKV 1386
Db 2240 PQVNNPREMLQVDFQTKMKVTGTTQGVKSLTSMYVKEFLISSSQDGHQWTLFFQNGKV 2299
QY 1387 KVFQGNQDSFTPVVNSLDBPLLTRYLRIHPQSHVQIALRMEVILGCEAODLY 1438
Db 2300 KVFQGNQDSFTPVVNSLDBPLLTRYLRIHPQSHVQIALRMEVILGCEAODLY 2351

RESULT 9
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor V
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10/132,829
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-829-4
```

```
Query Match 94.0%; Score 7227; DB 15; Length 2351;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY 1 ATRRYLGAVELSWDMQSDLCGLPVDARFPRVPRKSPSPNTSVVYKTLFVEFTVHLEN 60
Db 20 ATRRYLGAVELSWDMQSDLCGLPVDARFPRVPRKSPSPNTSVVYKTLFVEFTDHLFN 79
QY 61 IAKPRPPMGLLGPTIQAEVYDTVITLKNMASHVPSLHAGVSVYKASEGAEYDDQTSQ 120
Db 80 IAKPRPPMGLLGPTIQAEVYDTVITLKNMASHVPSLHAGVSVYKASEGAEYDDQTSQ 139
QY 121 REKEDDKVPFGSGHYTVQVLKENGPMASDPLCTYSYLSHVDLVKDNLSGILGALLVCR 180
Db 140 REKEDDKVPFGSGHYTVQVLKENGPMASDPLCTYSYLSHVDLVKDNLSGILGALLVCR 199
QY 181 EGSLAKEKTQTLHKFILLFAVDFECKSWHSETNSLMQDRDAASARAWPKMHTVNGYNR 240
Db 200 EGSLAKEKTQTLHKFILLFAVDFECKSWHSETNSLMQDRDAASARAWPKMHTVNGYNR 259
QY 241 SLPLGIGCHRSVYVHWIYGMGTTPVHSIFILEGHTFLVRNHRQASLEISPTITLTAOTLL 300
Db 260 SLPLGIGCHRSVYVHWIYGMGTTPVHSIFILEGHTFLVRNHRQASLEISPTITLTAOTLL 319
QY 301 MDLQGFLLFCHLSSHQHGMAYVYKDCSPEPQLRMKNNEAEYDDDLTDEMDVVRP 360
Db 320 MDLQGFLLFCHLSSHQHGMAYVYKDCSPEPQLRMKNNEAEYDDDLTDEMDVVRP 379
QY 361 DDNSPSPTQIRSVAKKHPKTVVHYIAAEEDWDYAPLVLAPEDDRSYKQYLNNGPQIRG 420
Db 380 DDNSPSPTQIRSVAKKHPKTVVHYIAAEEDWDYAPLVLAPEDDRSYKQYLNNGPQIRG 439
QY 421 RYKVRPMAYTDEFTKTRAIQHESGILGPLYGEVDTLLIIFKNOASRPNYIYPHGI 480
Db 440 RYKVRPMAYTDEFTKTRAIQHESGILGPLYGEVDTLLIIFKNOASRPNYIYPHGI 499
QY 481 TDVRLYSRRLPKGVKHLKDPPILEGRIKFKWTVVDEGDTKSDPRCLTIYYSFVYNE 540
Db 500 TDVRLYSRRLPKGVKHLKDPPILEGRIKFKWTVVDEGDTKSDPRCLTIYYSFVYNE 559
QY 541 RDLASGLIGPLLIYKESVDQGRNQIMSDKENVILFVSFEDNRNRSWYLTENTQRLPNPAG 600
```

Db 560 RDLASGLIGLLICYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFELNPAG 619
Qy 601 VQLEDEFOASNIWHSINGVVDLSQLSVCLHEVAYWYILSIGAQTDPLSVFSGYTPKH 660
Db 620 VQLEDEFOASNIWHSINGVVDLSQLSVCLHEVAYWYILSIGAQTDPLSVFSGYTPKH 679
Qy 661 KMYVEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNGMTALLKVSCKDKNTGDYIE 720
Db 680 KMYVEDTLTLPFPGSETVFMENPGLWILGCHNSDFRNGMTALLKVSCKDKNTGDYIE 739
Qy 721 DSYEDISAYLLSKNNAIEPRSP----- 742
Db 740 DSYEDISAYLLSKNNAIEPRSPSQNSRHPSTRQKQFNATTIPENDIEKTDPEFAHRTMP 799
Qy 743 ----- 742
Db 800 KIQNVSSDLLMLLRSPPHGLSLDLQAKYETPDDPSPGAIDSNNSLSMTHFRPQ 859
Qy 743 ----- 742
Db 860 LHESGDMVFPESGLQLRLNEKLGTTAATELKKLDFKVSSTNNLISTIPSDNLAAGTDN 919
Qy 743 ----- 742
Db 920 TSSLGPPMPVHYDSOLDTTLFGKXSPLTESGGPLSLENNDSKLLBSGLMNSQESSW 979
Qy 743 ----- 742
Db 980 GKNVSTESGRLPKGRAHGPALLTKDNALFKVISLTKNTKSNNSATNRKTHIDGPSL 1039
Qy 743 ----- 742
Db 1040 LIENSPVMQNILESDFKVTPLJHRLMLDKNATLRLNHSNKTSSKNMVMQOK 1099
Qy 743 ----- 742
Db 1100 KEGPIPPDAQNDMSFFKMLFUPESARWQTHGKNSLNSGGQSPKQVLVSLGPEKSVBG 1159
Qy 743 ----- 742
Db 1160 QNFLSEKNVVGKGFVKDGLKEMVFPSSRNLFNLNLHENTHNEKQKIQEIEIK 1219
Qy 743 ----- 742
Db 1220 KETLIQENVLPQIHTVTKNFMKNLFLSTRQNVGSYDGAAPVLQDPRSLNDSNR 1279
Qy 743 ----- 742
Db 1280 TKKHTAFSKGEBENLEGLGNQTKQIVKQYACTTRISNTSQNFVTQSRKALKQFRL 1339
Qy 743 ----- 742
Db 1340 PLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKEGALTQSPSLDCLFRSHI 1399
Qy 743 ----- 742
Db 1400 PQANRSPFIKAVSPSPSIRPIYLRVLFDQNSHLPAAVYRKDGSGVQESSHFLQAKK 1459
Qy 743 ----- 742
Db 1460 NNLSLAILTLEMTGDQREVGLSGTSATNSVTYKVENVTLPKDDLPTKSGKVELLPKVIH 1519
Qy 743 ----- 742
Db 1520 YQKDLFPETSTNGSPGHLDVBGLLQGTGEGAIKWNEARNPGKVPFLVATESSAKTPSK 1579
Qy 743 ----- 742
Db 1580 LLDPLADNHYGTQIPKEWKSQESPEKTAFFKDDTLISLNACBSNHAIAINEGQNK 1639
Qy 743 -----SQNPPVLKRRHQREITRTLLQSDQEEIDYDDTISVEMKEDFDIY 786
Db 1640 EIEVTWAKQGRTERLCSQNPVVLKRRHQREITRTLLQSDQEEIDYDDTISVEMKEDFDIY 1699

Qy 787 DEDEQSPSPFOKTRHPTFAAVERLWDYGMSSSHVLNRAQSGSVPOFKKVVFOEFTD 846
Db 1700 DEDEQSPSPFOKTRHPTFAAVERLWDYGMSSSHVLNRAQSGSVPOFKKVVFOEFTD 1759
Qy 847 GSFTQPLXRGELNEHLGLGPYIRAEVDNIWVTRNOASRPVSPYSSLSIYEEODROGA 906
Db 1760 GSFTQPLXRGELNEHLGLGPYIRAEVDNIWVTRNOASRPVSPYSSLSIYEEODROGA 1819
Qy 907 EPRKNFVKNETKTYFMKVOHEMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGPLVCHT 966
Db 1820 EPRKNFVKNETKTYFMKVOHEMAPTKDFCDKAWAYFSDVDLEKDVHSLGLIGPLVCHT 1879
Qy 967 NTLNPAHQROVTVQSFALFTTIFDETKSWYFTENMERNCRAPCNQMBDPTPKENYRFA 1026
Db 1880 NTLNPAHQROVTVQSFALFTTIFDETKSWYFTENMERNCRAPCNQMBDPTPKENYRFA 1939
Qy 1027 INGYIMDTLPLGLVMAQDQIRWVLLSMGSNENIHSIFSGHVFTVRKKKEYKALYNLYP 1086
Db 1940 INGYIMDTLPLGLVMAQDQIRWVLLSMGSNENIHSIFSGHVFTVRKKKEYKALYNLYP 1999
Qy 1087 GVFTVEMLPSKAGIWRVECLTGEHLHAGMSTLFLVYSNKKCTPLGMASGHTRDFOITAS 1146
Db 2000 GVFTVEMLPSKAGIWRVECLTGEHLHAGMSTLFLVYSNKKCTPLGMASGHTRDFOITAS 2059
Qy 1147 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQFSSLYISQ 1206
Db 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQFSSLYISQ 2119
Qy 1207 FIIMYSLDGKKWQTVRGNSTGTLMVFFGNVDSSGKIKHINFNPPIIARYIRLPHTHYSIRS 1266
Db 2120 FIIMYSLDGKKWQTVRGNSTGTLMVFFGNVDSSGKIKHINFNPPIIARYIRLPHTHYSIRS 2179
Qy 1267 TIRBELMGCDLNSCMPLGSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 1326
Db 2180 TIRBELMGCDLNSCMPLGSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 2239
Qy 1327 PQVNNPKEMLVQDFQTKMKVTGVTQGVKSLTSMYKKEFLISSQDGHQWTLFPQNGKV 1386
Db 2240 PQVNNPKEMLVQDFQTKMKVTGVTQGVKSLTSMYKKEFLISSQDGHQWTLFPQNGKV 2299
Qy 1387 KVFQGNQDSFTPVVNSLDPPLLTLYLRHPSQWVHOIALRMEVLGCEAQDLY 1438
Db 2300 KVFQGNQDSFTPVVNSLDPPLLTLYLRHPSQWVHOIALRMEVLGCEAQDLY 2351

RESULT 10

US-10-172-712-27
; Sequence 27, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.
; APPLICANT: GALE, ANDREW J.
; APPLICANT: GETZOFF, ELIZABETH D.
; APPLICANT: PELLEQUER, JEAN-LUC
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
; FILE REFERENCE: 4198-400US1
; CURRENT APPLICATION NUMBER: US/10/172,712
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 60/298,578
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-172-712-27

Query Match 94.0%; Score 7227; DB 15; Length 2351;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

Db 2180 TLRMELMGCDLNSCSPMLGMSKASDAQITASSYFTNMPATWSPSKARHLQGRSNAR 2239

Qy 1327 PQVNNPKELQVDFQKTKMTGVTTCQVKSLTSMYKKEFLISSQDGHQWTLFPQNGKV 1386

Db 2240 PQVNNPKELQVDFQKTKMTGVTTCQVKSLTSMYKKEFLISSQDGHQWTLFPQNGKV 2299

Qy 1387 KVFQGNODSFTPPVNSLDPPLTLRLYRHPQSWHQIALRMEVLGCEAQDLY 1438

Db 2300 KVFQGNODSFTPPVNSLDPPLTLRLYRHPQSWHQIALRMEVLGCEAQDLY 2351

RESULT 11

US-10-095-718-4

; Sequence 4, Application US/10095718

; Publication No. US20020131956A1

; GENERAL INFORMATION:

; APPLICANT: Walsh, Christopher

; APPLICANT: Chao, Hengjun

; APPLICANT: Burshtein, Haim

; APPLICANT: Lynch, Carmel

; APPLICANT: Stepan, Tony

; APPLICANT: Munson, Keith

; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and

; TITLE OF INVENTION: Methods of Using the Same

; FILE REFERENCE: 35052/204375

; CURRENT APPLICATION NUMBER: US/10/095,718

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 09/689,430

; PRIOR FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/158,780

; PRIOR FILING DATE: 1999-10-12

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 1431

; TYPE: PRT

; ORGANISM: canine B-domain deleted factor VIII

US-10-095-718-4

Query Match 85.8%; Score 6599; DB 14; Length 1431;

Best Local Similarity 85.6%; Pred. No. 0;

Matches 1229; Conservative 80; Mismatches 99; Indels 28; Gaps 6;

Qy 1 ATRRYLGAVELSDWYMQSD-LGELPVDARPPRPVPSKPPFTSVVYKTLFVFTVHLF 59

Db 20 ATRKYLGAVELSDWYMQSDLLSALHADTSFSRVPGLTITSVYKTVFVFTDDLF 79

Qy 60 NIAKPRPNWGLLPTIOAEVYDTVTITKQMAHPVSLHAGVSYWKAEGAEYDDQTS 119

Db 80 NIAKPRPNWGLLPTIOAEVYDTVTITKQMAHPVSLHAGVSYWKAEGAEYDDQTS 139

Qy 120 QREKEDKVPFGSGSHYVMQVLKXENGPMSADPLCLTYSYLSHVDLVKDLNSGLIGALLVC 179

Db 140 QKEKEDDNTVPGSSHYYVMQVLKXENGPMSADPLCLTYSYLSHVDLVKDLNSGLIGALLVC 199

Qy 180 REGSLAKEKQTQHLKFIPLPAVDEGKSHSETKNSLMQDRDAASARAFPMHTVANGYN 239

Db 200 KEGSLAKEKQTQHLKFIPLPAVDEGKSHSETNASLTQ-----AEAQHELATINGYN 253

Qy 240 RSLPGLIGCHRSVYWHVIGMTTPEVHSITFLEGHTFLVNRHQASLEISPTIFLTAQTL 299

Db 254 RSLPGLTVCHRSVYWHVIGMTTPEVHSITFLEGHTFLVNRHQASLEISPTIFLTAQTF 313

Qy 300 LMDLQGLFLPCHTSSHQDGMAYVYKVDSCFEFPQLMKNEBAEDYDDDLTDSMDVVR 359

Db 314 LMDLQGLFLPCHTSSHQDGMAYVYKVDSCFEFPQLMKNEBAEDYDDDLTDSMDVVR 372

Qy 360 FDDNDSFPTQIRSVAKGPKTWHYIAAREEDWDYAPLVAPDDRYSKYQVNNPQRI 419

Db 373 FDDNDSFPTQIRSVAKGPKTWHYIAAREEDWDYAPLVAPDDRYSKYQVNNPQRI 432

Qy 420 GRKYKVRFMAYTDETFKTRATQESGILGPLLYGEVGDTLIIIFKQASRPNIYPHG 479

RESULT 12

US-10-131-510A-39

; Sequence 39, Application US/10131510A

; Publication No. US20030166536A1

; GENERAL INFORMATION:

; APPLICANT: Lollar, John S

Db 433 GKXKVKRVFVAYTDETFKTRATQESGILGPLLYGEVGDTLIIIFKQASRPNIYPHG 492

Qy 480 ITDVRPLYSRLPKGVKHLKDXFPILPGEIPFKYKWTVTVEGPTKSDPRCLTRYSSFVNM 539

Db 493 INTVPLHTQRLPKGVKHLKDXFPILPGEIPFKYKWTVTVEGPTKSDPRCLTRYSSFINL 552

Qy 540 ERDLASGLIGPLATICYKESVDQEGNQITMSDKRNVILFSEVEDNRSWYLTENIQFLNPA 599

Db 553 ERDLASGLIGPLATICYKESVDQEGNQITMSDKRNVILFSEVEDNRSWYLTENIQFLNPA 612

Qy 600 GVQLEDPEFOASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSVFSGYTFK 659

Db 613 VVQPHDEPQLSNIMHSINGYVFDNLQLSVCLHEVAYWYILSIGAOTDPLSVFSGYTFK 672

Qy 660 HKWVYEDTTLFPFSGETVMSMENPGLWILGCHNSDFRNRMTALLKVSXCDKNTGDYY 719

Db 673 HKWVYEDTTLFPFSGETVMSMENPGLWILGCHNSDFRNRMTALLKVSXCDKNTGDYY 732

Qy 720 EDSYEDISAVLLSKNNAIPRSPSONPPVLKQHQREITRTLQSDQEEIDYDDTISVEMK 779

Db 733 EDTYEDIPPTLLNENVIKPRSFQN-----SRH-----PSTKBEK-----QLKMK 772

Qy 780 KEDFDYDEDENQSPRSFQKTKTRHYFIAAVERLDYGMSSSPHVLNRNRAQSGSVPOPKV 839

Db 773 REDFDYGVYENQGLRSFQKTKTRHYFIAAVERLDYGMSSSPHVLNRNRAQSGSVPOPKV 832

Qy 840 VFOEFTDGSFTQPLYRGELNEHLGLLGPYIRABVEDNIMVTFNRQASRPYSFYSLSIYE 899

Db 833 VFOEFTDGSFTQPLYRGELNEHLGLLGPYIRABVEDNIMVTFNRQASRPYSFYSLSIYE 892

Qy 900 EDQOAEPRKCNFVKPNETKTYFWKQHEMPTKDFDCAKAWFYSDVDLEKDVHSGILG 959

Db 893 EDEGQAEPRKCNFVKPNETKTYFWKQHEMPTKDFDCAKAWFYSDVDLEKDVHSGILG 952

Qy 960 PLAVCHTNTLNPAHGRQVTVQEPALFTTFIDETKSWYFTENMERNCRAPCNIQMEDPTEK 1019

Db 953 PLAVCHTNTLNPAHGRQVTVQEPALFTTFIDETKSWYFTENMERNCRAPCNIQMEDPTEK 1012

Qy 1020 ENYRFAHNGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIHSGHVFTVRKKEEYKM 1079

Db 1013 ENYRFAHNGYIMDTLPLGLVMAQDQIRIRWYLLSMGNSNENIHSIHSGHVFTVRKKEEYKM 1072

Qy 1080 ALYNYLPGVFETVEMLPSPKAGIWRVCECLIGELHAGMSTLELVYSNKKCOTPLGMASGHR 1139

Db 1073 AVNLNPGVFETVEMLPSPKAGIWRVCECLIGELHAGMSTLELVYSNKKCOTPLGMASGHR 1132

Qy 1140 DFQITASQGYQGWAPKARLHYSGSINAMSTKEPFSWIKVDLLAPMI IHGINTQAGARQF 1199

Db 1133 DFQITASQGYQGWAPKARLHYSGSINAMSTKEPFSWIKVDLLAPMI IHGINTQAGARQF 1192

Qy 1200 SLYISQFTIMYSLDQKXQTVRGNSGTGLAVFFGNVDSGGIKHNFNPITARIYRLHP 1259

Db 1193 SLYISQFTIMYSLDQKXQTVRGNSGTGLAVFFGNVDSGGIKHNFNPITARIYRLHP 1252

Qy 1260 THYSRSTRMELMGCDLNSCSPMLGMSKASDAQITASSYFTNMPATWSPSKARHLQ 1319

Db 1253 THYSRSTRMELMGCDLNSCSPMLGMSKASDAQITASSYFTNMPATWSPSKARHLQ 1312

Qy 1320 GRSNAWRPVNNPKELQVDFQKTKMTGVTTCQVKSLTSMYKKEFLISSQDGHQWTL 1379

Db 1313 GRSNAWRPVNNPKELQVDFQKTKMTGVTTCQVKSLTSMYKKEFLISSQDGHQWTL 1372

Qy 1380 EFQNGKVKVFCQGNODSFTPPVNSLDPPLTLRLYRHPQSWHQIALRMEVLGCEAQ 1435

Db 1373 EFQNGKVKVFCQGNODSFTPPVNSLDPPLTLRLYRHPQSWHQIALRMEVLGCEAQ 1428

```

; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 39
; LENGTH: 1443
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: factor VIII lacking
; OTHER INFORMATION: the B domain
US-10-131-510A-39

Query Match      84.4%; Score 6489; DB 12; Length 1443;
Best Local Similarity 83.4%; Pred. No. 0;
Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3;

QY      1  ATRRYLGAVELSDWYQSD-LGELPVDARPPRPVPSFPFNTSVVYKTLFVEFTVHLF 59
DB      20  ATRRYLGAVELSDWYQSELLRELVDTFPATAPGALPLGSLVKKYVFEFTDQLF 79

QY      60  NIAKPRPPNGLGPTTQAEYDVTWITLKNASHPVSLHAGVSYWKASGAEYDDQTS 119
DB      80  SVARPRPFNMLGPTTQAEYDVTWITLKNASHPVSLHAGVSYWKASGAEYEDHTS 139

QY      120  QREKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALVYC 179
DB      140  QREKEDKVLPGKSTYVQVLKENGPTASDPPCLTYSLSHVDLVKDLNSGLIGALLVC 199

QY      180  REGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASAPAWPKMHTVANGYN 239
DB      200  REGSLTRERTQNLHEFVLLFAVDEGKSWHSARNDSTWTRAMDPAAPAQAAMHTVANGYN 259

QY      240  RSLPGLIGCHRSYVHWVIGNGTTPYVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 299
DB      260  RSLPGLIGCHRSYVHWVIGNGTTPYVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTF 319

QY      300  LMDLQGLLFCISSHQDQMEAYKVVDSCPEEPQLRMKNNEBAEDYDDDLTDSMDVVR 359
DB      320  LMDLQGLLFCISSHHGGMEAHVRVESCAREPQLRKADER-EDYDDNLYSDMDVVR 378

QY      360  FDDNSPSTQIRSVAKKHPTWVHYTAABEEDWDYAPLVAPDDRYSYKSQYLNNGPQRI 419
DB      379  LDGDDVSFFQIRSVAKKHPTWVHYISABEEDWDYAPAVSPSDRSYKSQYLNNGPQRI 438

QY      420  GRKYKVRWAYTDETEKTRALCHESGILGPLYGVGDTLIIIPKQASRPYNYLPHG 479
DB      439  GRKYKARFVAYDVTETKTRALPYESGILGPLYGVGDTLIIIPKQASRPYNYLPHG 498

QY      480  ITDVRPLYSRLPKGVKHLKDFPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 539
DB      499  ITDVSALHPGKLLKGWHLKDMPLPGETFKYKWTVTVEGPTKSDPRCLTRYSSINL 558

QY      540  ERDLASGLIGFLPICKESVDQRGNQIMSKRNVILFSPVDENRWSWYLTENIQRFNPNA 599
DB      559  EKDLASGLIGFLPICKESVDQRGNQIMSKRNVILFSPVDENRQSWYLAENIQRFNPND 618

```

```

QY      600  GVQLBEPFOASNIMESINGYVFDLSQLSVCLHEVAWYILSIGAOTDFLSVFFSGYTFK 659
DB      619  GLQFQDPEFOASNIMESINGYVFDLSQLSVCLHEVAWYILSIGAOTDFLSVFFSGYTFK 678
QY      660  HKMVEDTLTLFPFSGETVFMSEMENPCLMILGCHNSDFRNRGTMALLKVSXCDKATGDYY 719
DB      679  HKMVEDTLTLFPFSGETVFMSEMENPCLMILGCHNSDFRNRGTMALLKVSXCDKATGDYY 738
QY      720  ZDSYEDISAYLLSKNNAIBRPSQNPVPLKQRETRITLQSDDEEDYDDTISVEMK 779
DB      739  DNTYEDIPGLLSGKNVIEP-----RDISLPTFPQBEDKMDYDDIFSTETK 784
QY      780  KEDFDIYDEDENOSPRSFQKTRHYFIAAVERLWDYGMSSSPHLVRNRAQSGSVPOPKV 839
DB      785  GEDFDIYGEDENQDPRSFQKTRHYFIAAVERLWDYGMSSSPHLVRNRAQSGSVPOPKV 844
QY      840  VFOEPTGSGTQPLYRGSLEHGLGPIYRAEVEDNIMVTFRQASRPYSFYSSLSIYE 899
DB      845  VREFADGSGTQPSYRGELNKHGLGPIYRAEVEDNIMVTFRQASRPYSFYSSLSIYP 904
QY      900  EDQOQAEPRKVFKNETKTYFWKHMAPTKDFDCKAWAYFSDVDLEKDVHSGLIC 959
DB      905  DDQOQAEPRHNFVQNETRTYFWKHMAPTKDFDCKAWAYFSDVDLEKDVHSGLIC 964
QY      960  PLIVCHTNTINPAHGRQVTVQSFALFETTFDETKSWYFTENMERNCRAPCNIQMEDPTEK 1019
DB      965  PLIVCHTNTINPAHGRQVTVQSFALFETTFDETKSWYFTENMERNCRAPCNIQMEDPTEK 1024
QY      1020  ENTRFHAINGYVMDTLPLGLVMAQNRIRWYLLSMGNSNENIHSIHPSHGVFVRKKEEYKM 1079
DB      1025  ENTRFHAINGYVMDTLPLGLVMAQNRIRWYLLSMGNSNENIHSIHPSHGVFVRKKEEYKM 1084
QY      1080  ALVNLVPGVFETVEMLPKSGAGIWRVCLIGEHLAGHAGHSTFLVYSNKQCTPLGMASCHIR 1139
DB      1085  AVNLVPGVFETVEMLPKSGAGIWRVCLIGEHLAGHAGHSTFLVYSNKQCTPLGMASCHIR 1144
QY      1140  DFOITASGOYQWAPKLABLHYSGSINASTKEPFSWIKVLDLAPMIIEGKITQGARQKF 1199
DB      1145  DFOITASGOYQWAPKLABLHYSGSINASTKEPFSWIKVLDLAPMIIEGKITQGARQKF 1204
QY      1200  SSIYISQFTIMVSLDQKKQTVRGNSGTGLMVFVGVNDSSGKGNIFNPPIIARYIRLHP 1259
DB      1205  SSIYISQFTIMVSLDQKKQTVRGNSGTGLMVFVGVNDSSGKGNIFNPPIIARYIRLHP 1264
QY      1260  THYSIRSTRMELMGCDLNSCMPLGMSKALSDAOITASSYFTNMFATWSPSKARHLQ 1319
DB      1265  THYSIRSTRMELMGCDLNSCMPLGMSKALSDAOITASSYFTNMFATWSPSKARHLQ 1324
QY      1320  GRSNAWRPOVNNPKQLVDFOKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTL 1379
DB      1325  GRTNAWRPVSAEELQVDLQTKVKTGTTTQGVKSLSSMYVKEFLVSSSQDGRWTL 1384
QY      1380  FFQNGKVKVFCQGNQDSFTPVNSLOPPLLTLYLRHPSQWVHQTALRMEVLGEADLY 1438
DB      1385  FLDGHTKVFQGNQDSFTPVNSLOPPLLTLYLRHPSQWVHQTALRMEVLGEADLY 1443

```

RESULT 13

```

US-10-187-319-39
; Sequence 39, Application US/10187319
; Publication No. US20030068785A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESS: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/187,319
 FILING DATE: 27-Aug-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/523,656
 FILING DATE: 2000-03-10
 APPLICATION NUMBER: US 09/037,601
 FILING DATE: 1998-03-10
 APPLICATION NUMBER: WO PCT/US97/11155
 FILING DATE: 1997-06-26
 APPLICATION NUMBER: US 08/670,707
 FILING DATE: 1996-06-26
 ATTORNEY/AGENT INFORMATION:
 NAME: Greenlee, Lorraine L.
 REGISTRATION NUMBER: 27,894
 REFERENCE/DOCKET NUMBER: 75-95K
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 303/499-8080
 TELEFAX: 303/499-8089
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1443 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-10-187-319-39

Query Match 84.4%; Score 6489; DB 15; Length 1443;
 Best Local Similarity 83.4%; Pred. No. 0;
 Matches 1200; Conservative 104; Mismatches 119; Indels 16; Gaps 3;

1 ATRRYVIGAVELSDWYQSD-LGELPVDARPPRPVPSFPFNTSVVYKTLFVEFTVHLF 59
 20 AIRRYVIGAVELSDWYQSDRELHVDTRFATGALPLGPSVLYKTVFVEFTDQLF 79
 60 NIAKPRPWWGLIGPTQARVYTVITLQMAHPVSLHAGVSWKASGAEYDDQIS 119
 80 SVARPRPWWGLIGPTQAEVYTVVITLQMAHPVSLHAGVSWKASGAEYDDHTS 139
 120 QREKEDKVPFGSGSHYVVOVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC 179
 140 QREKEDKVLPGKSQTVVQVLKENGFTASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC 199
 180 REGSLAKETQTLHKFTLLPAVDEGKSWHSETKNSLMQDRDASAPAWPKMHTVNGYN 239
 200 REGSLTRERTQNLHEFVLLPAVDEGKSWHSEARNDSTWAMODAPAPAPAMHTVNGYN 259
 240 RSLPGLIGCHRKSVYWHVIGMGTTPVHSTPLRGHTFLVNHROASLEISPTFELTAQTL 299
 260 RSLPGLIGCHRKSVYWHVIGMGTTPVHSTPLRGHTFLVNHROASLEISPLTFLTAQTF 319
 300 LMDLGQFLFCHTSSHQHGMAYVYKVDSPCEPQRLMKNNKNEAEYDDDLTDSEMDVVR 359
 320 LMDLGQFLFCHTSSHHGGMAYVYKVDSPCEPQRLMKNNKNEAEYDDDLTDSEMDVVR 378
 360 FDDNDSPTQIRSVAKGPKTWVHYIAEEEDWDYAPVLAPDDRSYKSYQVYNNPQR 419
 379 LDGDDVSPFTQIRSVAKGPKTWVHYISAEEDWDYAPVSPDSRSYKSLYNSGPKRI 438
 420 GRKYKVRMAYTDFPKTEATQHSGLIGPLLYGEVGDTLIIIPKNOASRPYNYIPHG 479
 439 GRKYKARFVAYTDVTFKTKKALPYESGILGPLLYGEVGDTLIIIPKNOASRPYNYIPHG 498
 480 ITDVRPLRRLPKGVKHLKDFPLPGEIFPKYKWTVTVEDGPKTSPRCLTRYSSFVNM 539
 499 ITDVSALHFGRLKGVKHLKDFPLPGEIFPKYKWTVTVEDGPKTSPRCLTRYSSINL 558

540 ERDLASGLIGPLLYCYKESVDQGRNQIMSDKRNVLFSVDEENRSHWLTENIORFLNPA 599
 559 EKDLASGLIGPLLYCYKESVDQGRNQIMSDKRNVLFSVDEENRSHWLTENIORFLNPD 618
 600 GVQLEDPFQASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGATQDFLSVFFSGYTFK 659
 619 GLQPDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYLSIGATQDFLSVFFSGYTFK 678
 660 HKMVEDTTLTPPFSGETVFMENPGLMILGCHNSDFRNRGTMALLKYSSCDKNTGDYY 719
 679 HKMVEDTTLTPPFSGETVFMENPGLMILGCHNSDFRNRGTMALLKYSSCDKNTGDYY 738
 720 EDSYEDISAVLLSKNNAIBRPSFQNPVLRKHQREITRTTLQSDOEEDYDDTISVEWK 779
 739 DNYIEDIPGLLSGQVIEP-----RDISLTPQPEESDKMDYDIDFISTETK 784
 780 KEDFDIYDEDENQSPRSFQKTRHYFIAAVERIMDYGMSSSPHVLNRRAQSGSVPOKKV 839
 785 GEDFDIYDEDENQSPRSFQKTRHYFIAAVERIMDYGMSSSPHVLNRRAQSGSVPOKKV 844
 840 VQCEFTDGSFTQPLRGELNEHLGLLGPYIRAEVEDNIMVTRNQASRPYSFYSSLSIZE 899
 845 VRFEPADGSFTQPSYRGELNKLGLLGPYIRAEVEDNIMVTRNQASRPYSFYSSLSIZE 904
 900 EDQOQABPRKPFVKNETKTYFWKQYOHMAPTKDFCDCKAWAYFSDVLEKDVHSGLIG 959
 905 DQOQOABPRKPFVKNETKTYFWKQYOHMAPTKDFCDCKAWAYFSDVLEKDVHSGLIG 964
 960 PLLVCHTNTLPAHGQVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNQMEDPTK 1019
 965 PLLICRANTLNAAGKQVTVQBFALFFTFIDETKSWYFTENMERNCRAPCNQMEDPTK 1024
 1020 ENYRFAINGYIMDLPLGLVMAQDQIRIYLLSMGSENIHSHFSGHVTYVRKKEEYKM 1079
 1025 ENYRFAINGYIMDLPLGLVMAQDQIRIYLLSMGSENIHSHFSGHVTYVRKKEEYKM 1084
 1080 ALYNYLPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKKCOTELGMASSGR 1139
 1085 AVYNYLPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKKCOTELGMASSGR 1144
 1140 DFQITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGKIQGARQKF 1199
 1145 DFQITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGKIQGARQKF 1204
 1200 SSIYISQFTIMYSLDGKQWOTYRGSTGTLPFGNVDSGSKHNFNPPIARIRLHP 1259
 1205 SSIYISQFTIMYSLDGKQWOTYRGSTGTLPFGNVDSGSKHNFNPPIARIRLHP 1264
 1260 THYSIRSTLRBELMGCDLNSCMPLGMSKASDAQITASSYFTNMFMATWSPSKARLHLQ 1319
 1265 THYSIRSTLRBELMGCDLNSCMPLGMSKASDAQITASSYFTNMFMATWSPSKARLHLQ 1324
 1320 GRSNARVQNNPKWQLQVDFOKTKVGTVTQGVKSLTSMYKFEKLSSQDGHQWTL 1379
 1325 GRTNARVRSABEWLQVDFOKTKVGTVTQGVKSLTSMYKFEKLSSQDGHQWTL 1384
 1380 RFQNGKVKVFCQGNQDSFTPVNNSLDPULLTRVLRTHPOSVHOIARLMEVLGCEAQDLY 1438
 1385 FLQDGHKTVFQGNQDSFTPVNNSLDPULLTRVLRTHPOSVHOIARLMEVLGCEAQDLY 1443

RESULT 14

US-10-131-510A-37
 ; Sequence 37, Application US/10131510A
 ; Publication No. US20030166536A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S
 ; TITLE OF INVENTION: Modified Factor VIII
 ; FILE REFERENCE: 75-95J
 ; CURRENT APPLICATION NUMBER: US/10/131,510A
 ; CURRENT FILING DATE: 2002-07-10
 ; PRIOR APPLICATION NUMBER: U.S. 09/315,179
 ; PRIOR FILING DATE: 1999-05-20

; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2133
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-131-510A-37

Query Match 81.0%; Score 6227; DB 12; Length 2133;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

QY 1 ATRRYVLGAVELSDWYMSD-IGELPYDARPPRPVPSFPFNTSVVKKTLFVFTVHLF 59
DB 20 ATRRYVLGAVELSDWYMSD-IGELPYDARPPRPVPSFPFNTSVVKKTLFVFTVHLF 79
QY 60 NIARPPPMGLGPTTQABYDVTVITLNKMAHPVSLHAGVSVYKASEGAEYDDQTS 119
DB 80 SVARPPPMGLGPTTQABYDVTVITLNKMAHPVSLHAGVSVYKASEGAEYDDQTS 139
QY 120 QRKEDDKVPFGSGSHYVWVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC 179
DB 140 QRKEDDKVLPFGSKQTVVWVLKENGPMASDPLCLTVSYLSHVDLVKDLNSGLIGALLVC 199
QY 180 REGSLAKEKOTLHKFTLLFAVDEGKSWHSEKNSLMQDRDAASARAPPMHTVNGYN 239
DB 200 REGSLAKEKOTLHKFTLLFAVDEGKSWHSEKNSLMQDRDAASARAPPMHTVNGYN 259
QY 240 RSLPGLIGCHRSYVHVIWGTTPVHSIFLECHTFLVNRHQAASLEISPIFLTAQTL 299
DB 260 RSLPGLIGCHRSYVHVIWGTTPVHSIFLECHTFLVNRHQAASLEISPIFLTAQTL 319
QY 300 LMDLQFLPCHLSSHQDCHAEVYKVDSCPEPFLKMNKNEAEYDDDLTDSMDVVR 359
DB 320 LMDLQFLPCHLSSHQDCHAEVYKVDSCPEPFLKMNKNEAEYDDDLTDSMDVVR 378
QY 360 FDDNPSFQIRSVAKKPKTWVHYTAAEEDWDYAPLVLPDDRYSKYSLYLNNGPQRI 419
DB 379 LDGDDVSPFQIRSVAKKPKTWVHYTAAEEDWDYAPLVLPDDRYSKYSLYLNNGPQRI 438
QY 420 GRKYKVRWYATDFTKTRATQHSGLIGPLLYGEVGTLLIIIFKNQASRPYNTYPHG 479
DB 439 GRKYKVRWYATDFTKTRATQHSGLIGPLLYGEVGTLLIIIFKNQASRPYNTYPHG 498
QY 480 ITDVRPLYSRLPKVGHKLDFTLPCIEIFKYKWTYVDEGPKSDPRCLTRYSSFVNM 539
DB 499 ITDVRPLYSRLPKVGHKLDFTLPCIEIFKYKWTYVDEGPKSDPRCLTRYSSFVNM 558
QY 540 ERDLASGLIGPLLYCYKESYDQGNQIMSDKRVILFSVFEDNRSWYLTENIORFLPNEA 599
DB 559 ERDLASGLIGPLLYCYKESYDQGNQIMSDKRVILFSVFEDNRSWYLTENIORFLPNEA 618
QY 600 GVQLEDPEFOASIMHSINGYVDSLSVCLHEVAYWYILSICAOITDFLVSFSGYTK 659
DB 619 GVQLEDPEFOASIMHSINGYVDSLSVCLHEVAYWYILSICAOITDFLVSFSGYTK 678
QY 660 HKMVIYEDTLTLPFSGETVPMENPGLMTLGCNDSDFNRGHTALLKYSSCDKNTGDY 719
DB 679 HKMVIYEDTLTLPFSGETVPMENPGLMTLGCNDSDFNRGHTALLKYSSCDKNTGDY 738
QY 720 EDSYEDISAYLLSKNNAIEPRFSQN----- 745

DB 739 DNTYEDIPGFLSGKNIPIPRPSAQNSQKQFQITSPEDDVELDPOSGERTQAL 798
QY 746 ----- 745
DB 799 BELSVPSGDSMLLQGNPAPHGSSSDLOEARNADDYLPGARERTWPAASAAARLPELH 858
QY 746 ----- 745
DB 859 HSAERVLTPEPEKELKLDKSKSSSDLLKTSFTIPSDTLSAERTHTSLGPHHPQVNR 918
QY 746 ----- 745
DB 919 SOLGALVLGKNSHFTGACVPLGSTEDHSSGENVSPESDGIPEKERAHPASLTKD 978
QY 746 ----- 745
DB 979 DVLKVNISLVKTNKARVYLKTNRKIHIDDAALTENRASATFMDKNKTASGLNHVSNWI 1038
QY 746 ----- 745
DB 1039 KGPLGKNPLSSRSGPPELLTSSGSGKSVKQSGQGRIRVAVEBEELSGKEMMLPNS 1098
QY 746 ----- 745
DB 1099 LTFLTNSADVQGNTHSQKKSREMERREKLVBQKVDLPQVYTATGATGNFLRNIHQST 1158
QY 746 ----- 745
DB 1159 EPSVEGPDGSHAPVQDQSRSLNDSAEAEATHIAHFAIREAEAPLEAPGNRTGPGPSAV 1218
QY 746 ----- 745
DB 1219 PRVVKSLQIRLPLEEIKPERGVNLNATSRWSBSSPILQAGKRNLSLPFLTLEMAGG 1278
QY 746 ----- 745
DB 1279 QGKISALGKSAAGPLASGLKAVLSSAGLSEASGKAEFLPKVRVHREDLLPQKTSNVSC 1338
QY 746 ----- 745
DB 1339 AHGDLQGEIFLQKTRGPVNLKNVNRPGRTPSKLLGPPMPKWESELESKSPKALTAKDII 1398
QY 746 ----- 745
DB 1399 SLPLDRHESNHSIAAKNEQAEQAEAAWTQKGGPGLCAPKPPVLRHRQRDISLPTFQ 1458
QY 764 DQSEIDYDDTISVENKKEDEFDIYDSDENOSPEFSOKTTHYETAAVERLWYGMSSSHV 823
DB 1459 ESKKMDYDDIFSTETKGEFDIYDSDENODPSRQKTRHYFIAAVEQQLWYGMSESPPA 1518
QY 824 LBNRAQSGSVPOFKKVVFOEFTDGSFTQPLYXGELNEHLGLLGPYIRAEVEDNIMVTFEN 883
DB 1519 LBNRAQSGSVPOFKKVVFOEFTDGSFTQPLYXGELNEHLGLLGPYIRAEVEDNIMVTFEN 1578
QY 884 QASRPYSFVSSLSIYBDEORQABPRKOFVKNETKTFWKVQHMAPTKDDFCCKAWAY 943
DB 1579 QASRPYSFVSSLSIYBDEORQABPRKOFVKNETKTFWKVQHMAPTKDDFCCKAWAY 1638
QY 944 FSDVLEKDVHSLGLLPLLVCHTNTLNPAGHQVTVQEFALPFTTIFDETKSWYFTENMER 1003
DB 1639 FSDVLEKDVHSLGLLPLLVCHTNTLNPAGHQVTVQEFALPFTTIFDETKSWYFTENMER 1698
QY 1004 NCRAPCNTQMEDPTPKENYRHAINGYIMDLPLGLVMAQDQRIWYLLSMGSENENIHSIH 1063
DB 1699 NCRAPCHLQMEDPTPKENYRHAINGYIMDLPLGLVMAQDQRIWYLLSMGSENENIHSIH 1758
QY 1064 FSGHVTVRKKEEYKALNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVY 1123
DB 1759 FSGHVTVRKKEEYKALNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVY 1818
QY 1124 SNKCOPTLGMASGHTRDFOITASGGYQGWAPKALRHYSGSNINAKSTKEPESWIKVDLLA 1183

Db 1039 KGPLGNPLSSERGPSPELLTSSGSGKSVKQSSGQGRIRVAVBEELSKEKEMLPNSE 1098
QY 746 ----- 745
Db 1099 LTFLTNSADYQGNTHSQGKSRBEMERREKLVEKYVDLPQVYTATGKPLRNIHQST 1158
QY 746 ----- 745
Db 1159 EPSVEGFDGSHAPVQDPSRLNDSABRAETHIAHFAIREAPLEAPGNRTGPPRSV 1218
QY 746 ----- 745
Db 1219 PRRVKQSLKQIRLPLEBEIKPERGVVLTATSTRWSESSPILQGAKNLNSLPFITLEMAGG 1278
QY 746 ----- 745
Db 1279 QGKISALGKSAAGPLASGLEKAVLSSAGLSEASGKAEPLPKVRVHREDLLPQKTSNVSC 1338
QY 746 ----- 745
Db 1339 AHGDLQEIIFLQKTRGPVNLKNVRPGRTPSKLGPPMPKWBSELSKPSKSTALRTKDI 1398
QY 746 ----- PPVLRHOREITRTTLOS 763
Db 1399 SLPLDRHESNHSIAAKNEGQAEQREAAWTKQGGPGLCAPKPPVLRHQHORDISLPTQP 1458
QY 764 DOBEIDYDDITISVEMKEDDIDYDEDENQSPRSQKTRHYFIAAVERLWDYGNSSSPHV 823
Db 1459 EEDKMDYDDIFSTETKGEDFDIYGEDENQDPRSFQKTRHYFIAAVERLWDYGNSSSPRA 1518
QY 824 LRNRQSGSVPOPKVVFQFTDGSFTQPLRYGELNEHLGLGPLYRAEVEDNIMTFRN 883
Db 1519 LRNRQSGSVPOPKVVFQFTDGSFTQPLRYGELNEHLGLGPLYRAEVEDNIMTFRN 1578
QY 884 QASRPYSFSSLSISYERDQGAEPKRFVKNPNETTYFWKVQHMAPTDEEDCKAWAY 943
Db 1579 QASRPYSFSSLSISYERDQGAEPKRFVKNPNETTYFWKVQHMAPTDEEDCKAWAY 1638
QY 944 FSDVDLEKDVHSLGIGLILVCHNTNTPAHGRQVTVQEFALPTTIFDETKSWYFTENMER 1003
Db 1639 FSDVDLEKDVHSLGIGLILVCHNTNTPAHGRQVTVQEFALPTTIFDETKSWYFTENMER 1698
QY 1004 NCRAPCNIQMEDPTFKENYRPHAINGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIH 1063
Db 1699 NCRAPCHLQNEDEFLKENYRPHAINGYIMDTPLGLVMAQDQIRWYLLSMGNSNENIHSIH 1758
QY 1064 FSGHVFTRKKEEYKMAVNLXPGVFETVEMLPKAGIMRVECLIGEHLHAGMSTLFLVY 1123
Db 1759 FSGHVFTRKKEEYKMAVNLXPGVFETVEMLPKAGIMRVECLIGEHLHAGMSTLFLVY 1818
QY 1124 SNKCOPLGNASGHIRDFOITASGOYQWAPKLARLHYSGSINAWSTKEPPSWIKUDLLA 1183
Db 1819 SKCOAPLGNASGHIRDFOITASGOYQWAPKLARLHYSGSINAWSTKEPPSWIKUDLLA 1878
QY 1184 PMIHGINTQGAQKFSLSYISQFIIMYSLDGRNQSYRGNSTGLVFFGNVDSSGKIH 1243
Db 1879 PMIHGINTQGAQKFSLSYISQFIIMYSLDGRNQSYRGNSTGLVFFGNVDSSGKIH 1938
QY 1244 NIFNPPIIARIYIRLHPHYSIRSLRMLGCDLNCMPMGESKAISDAQITASSYFT 1303
Db 1939 NIFNPPIIARIYIRLHPHYSIRSLRMLGCDLNCMPMGESKAISDAQITASSYFT 1998
QY 1304 NMEATSPSKARLHLCQNSNAWPOVNNPKEMLOVDPOKTKMYTGVTTQGVKSLTSMYV 1363
Db 1999 NIFATWSPSKARLHLCQNSNAWPOVNNPKEMLOVDPOKTKMYTGVTTQGVKSLTSMYV 2058
QY 1364 KEPL1SSQDGHQWTLFFQNGKVKVFGQNDQSFPPVNSLDPPLLTTRYLRHQPSSWHQI 1423
Db 2059 KEPLVSSQDGRWTLFLQDGHTKVFGQNDQSFPPVNSLDPPLLTTRYLRHQPSSWHQI 2118
QY 1424 ALRMEVLGCEAQDLY 1438
Db 2119 ALRMEVLGCEAQDLY 2133

Search completed: December 9, 2003, 17:09:44
Job time : 80 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 17:02:15 ; Search time 28 Seconds
(without alignments)
4938.949 Million cell updates/sec

Title: US-10-047-257-1

Perfect score: 7691

Sequence: 1 ATRRYLGAVELSWDYQSD.....VVHQIALRMEVLGCEAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:**

1: piri:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7227	94.0	2351	1 EZHU	coagulation factor
2	6227	81.0	2133	2 T42763	coagulation factor
3	6192	80.5	2319	2 A47004	coagulation factor
4	2390.5	31.1	2224	1 KFB05	coagulation factor
5	2372	30.8	2211	1 KFB05	coagulation factor
6	2369	30.8	2183	2 T42764	coagulation factor
7	1731.5	22.5	1069	1 KUHU	ferroxidase (BC 1.
8	1687.5	21.9	1059	1 A35210	ferroxidase (BC 1.
9	1104	14.4	216	2 A44258	factor VIII-associ
10	663	8.6	427	2 JC4915	ags protein precu
11	657	8.5	463	1 A36479	milk fat globule m
12	650	8.5	409	2 T11743	pp47 protein - pig
13	635	8.3	401	2 S65138	glycoprotein anti
14	635	8.3	427	2 S74211	PAS-6/7 protein pr
15	543	7.1	869	2 A25945	coagulation factor
16	443	5.8	927	1 JQ0948	A5 antigen precurs
17	424.5	5.5	218	2 A47285	milk fat globule p
18	306.5	4.0	3133	2 J52093	hemocytin - silkw
19	262	3.4	845	2 JC5256	adipocyte transcr
20	221	2.9	719	2 S51739	transcription repr
21	211	2.7	1283	2 T13799	neurexin IV - fru
22	207.5	2.7	737	2 T31349	hypothetical prote
23	206.5	2.7	737	2 T15615	hypothetical prote
24	193	2.5	913	2 A48280	receptor tyrosine
25	192	2.5	876	2 A49508	protein-tyrosine k
26	185.5	2.4	855	2 S42621	protein-tyrosine k
27	185.5	2.4	910	2 A53137	tyrosine kinase re
28	173.5	2.3	819	2 I48859	tyro 10 receptor k
29	171	2.2	1381	2 T31083	paranodin - rat

RESULT 1

EZHU

coagulation factor VIII precursor [validated] - human

N;Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant com

C;Species: Homo sapiens (man)

C;Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000

C;Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B

Hm. Mol. Genet. 1, 199-200, 1992

R;Gitschier, J.; Wood, W.I.

A;Title: Sequence of the exon-containing regions of the human factor VIII gene.

A;Reference number: I54318; MUID:93265012; PMID:1303178

A;Accession: I54318

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1921,'S',1921-2351 <RES>

A;Cross-references: GB:M8648; NID:g182381; PIDN:AAA52420.1; PID:g182383

R;Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb

Nature 312, 330-337, 1984

A;Title: Expression of active human factor VIII from recombinant DNA clones.

A;Reference number: A00525; MUID:85061548; PMID:6438526

A;Accession: A00525

A;Molecule type: mRNA

A;Residues: 1-2351 <WOO>

A;Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179

R;Tooie, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.B.; Pittman, D.D.

S, D.N.; Hewick, R.M.

Nature 312, 342-347, 1984

A;Title: Molecular cloning of a cDNA encoding human antihemophilic factor.

A;Reference number: I58059; MUID:85061550; PMID:6438528

A;Accession: I58059

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-74,'V',76-1259,'E',1261-2351 <RE2>

A;Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803

R;Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K

B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; J

DNA 4, 333-349, 1985

A;Title: Characterization of the polypeptide composition of human factor VIII:C and th

A;Reference number: A23584; MUID:86081164; PMID:3935400

A;Accession: A23584

A;Molecule type: mRNA

A;Residues: 1-2351 <TRU>

A;Cross-references: GB:M44113; NID:g182817; PIDN:AAA52485.1; PID:g182818

R;Ratton, D.; Rodriguez, H.; Vohar, G.A.

Biochemistry 25, 505-512, 1986

A;Title: Proteolytic processing of human factor VIII. Correlation of specific cleavage

ity.

A;Reference number: A26174; MUID:86159740; PMID:3082357

A;Accession: A26174

A;Molecule type: protein

A;Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <EAT

R;Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

neurexin IV - mous
L-ascorbate oxidas
hypothetical prote
probable multicopp
botulinum neurotox
toxin, nontoxic co
mxg protein - Bac
RNA-directed RNA p
nontoxic-nonhemag
probable laccase (
preprotein translo
hypothetical prote
probable copper ox
hypothetical prote
L-ascorbate oxidas

ALIGNMENTS

30 171 2.2 1385 2 T14158
31 149.5 1.9 578 2 S66353
32 149 1.9 1883 2 G82875
33 148 1.9 622 2 S62580
34 141.5 1.8 1196 2 S46430
35 140.5 1.8 1196 2 JQ1467
36 140.5 1.8 1217 2 T18209
37 139.5 1.8 2925 2 T00133
38 139 1.8 1193 2 JC4901
39 137.5 1.8 580 2 F84828
40 137.5 1.8 903 2 F82080
41 136.5 1.8 2657 2 T18497
42 135 1.8 343 2 T35030
43 134.5 1.7 791 2 T18031
44 134.5 1.7 1132 2 H82887
45 133.5 1.7 567 2 T44928

Biochemistry 31, 3315-3325, 1992
 A;Title: Identification and functional importance of tyrosine sulfate residues within re
 A;Reference number: A42348; MUID:92207952; PMID:1554716
 A;Accession: A42348
 A;Molecule type: protein
 A;Residues: 20-36;336-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-17
 A;Experimental source: recombinant material from Chinese hamster ovary cells
 A;Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
 R;Fay, P.J.; Smudzin, T.M.
 J. Biol. Chem. 264, 14005-14010, 1989
 A;Title: Intersubunit fluorescence energy transfer in human factor VIII.
 A;Reference number: A43986; MUID:89340500; PMID:2503509
 A;Accession: A43986
 A;Molecule type: protein
 R;Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
 J. Biol. Chem. 266, 740-745, 1991
 A;Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
 A;Reference number: A56109; MUID:91093266; PMID:1898735
 R;Gitschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
 Nature 312, 326-330, 1984
 A;Title: Characterization of the human factor VIII gene.
 A;Reference number: A56196; MUID:85061547; PMID:6438525
 A;Contents: annotation; introns
 R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
 Protein Sci. 4, 740-746, 1995
 A;Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o
 A;Reference number: A56216; MUID:9338127; PMID:7613471
 A;Contents: annotation; disulfide bonds
 A;Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
 R;Kjalkre, M.; Hedner, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
 Eur. J. Biochem. 234, 773-779, 1995
 A;Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A;Reference number: S63527; MUID:96163459; PMID:8575434
 A;Accession: S63527
 A;Molecule type: protein
 A;Residues: 733-752;753-759 <KJA>
 R;Lind, P.; Larsson, K.; Spira, J.; Sydow-Baekman, M.; Almetedt, A.; Gray, E.; Sandberg
 Eur. J. Biochem. 232, 19-27, 1995
 A;Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
 A;Reference number: S66445; MUID:96048024; PMID:7556150
 A;Accession: S66445
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 1668-1685 <LIN>
 C;Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
 C;Genetics:
 A;Gene: GDB:P9C
 A;Cross-references: GDB:119124; OMIM:306700
 A;Map position: Xq28-Xq28
 A;Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
 C;Function:
 A;Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
 A;Pathway: blood coagulation
 C;Superfamily: coagulation factor VIII; discoidein I amino-terminal homology; ferroxidase
 C;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilin A; pla
 F;1-19/Domain: signal sequence #scatus predicted <SIG>
 F;20-2351/Product: coagulation factor VIII #status experimental <MAT>
 F;20-740/Product: coagulation factor VIII heavy chain #status experimental <ACH>
 F;20-356/Domain: A1 <DAL>
 F;23-348/Domain: ferroxidase repeat homology <FO1>
 F;392-759/Domain: A2 <DA2>
 F;402-730/Domain: ferroxidase repeat homology <FO2>
 F;760-1667/Domain: B <DB0>
 F;1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
 F;1709-2038/Domain: A3 <DA3>
 F;1716-2038/Domain: ferroxidase repeat homology <FO3>
 F;2039-2191/Domain: C1 <DC1>
 F;2039-2188/Domain: discoidein I amino-terminal homology <DNI>
 F;2192-2351/Domain: C2 <DC2>
 F;2192-2345/Domain: discoidein I amino-terminal homology <DN2>
 F;60,258,601,776,803,847,919,962,982,1024,1074,1085,1204,1278,1301,1319,1403,1

F;172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #statu
 F;355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F;365-737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experimental
 F;391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F;414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
 F;759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F;1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
 F;1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F;1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F;2193-2345/Disulfide bonds: #status predicted
 Query Match 94.0%; Score 7227; DB 1; Length 2351;
 Best Local Similarity 61.6%; Pred. No. 0;
 Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;
 QY 1 ATRRYVLGAVELSDWYMQSDLGELVDARFPRVPKSPPPNTSVVYKTLFVEFTVHLEN 60
 DB 20 ATRRYVLGAVELSDWYMQSDLGELVDARFPRVPKSPPPNTSVVYKTLFVEFTVHLEN 79
 QY 61 IAKPRPPMGLLGPITQAEVYTVITLKNMASHPVSLHVGVSVMKASEGAYDDQTSQ 120
 DB 80 IAKPRPPMGLLGPITQAEVYTVITLKNMASHPVSLHVGVSVMKASEGAYDDQTSQ 139
 QY 121 RKEDDKVPPGSHYTVMOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
 DB 140 REKEDDKVPPGSHYTVMOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
 QY 181 EGS�AKEKTOTLHKETILLFAVDECKSWHSEPKNSLMQSDRAASARAPKMTVYGNR 240
 DB 200 EGS�AKEKTOTLHKETILLFAVDECKSWHSEPKNSLMQSDRAASARAPKMTVYGNR 259
 QY 241 SLPGLIGCHRKSVYVHVGMTTTPVHSIFLGHFTFLVNHRQASLEISPIFLTAQTLL 300
 DB 260 SLPGLIGCHRKSVYVHVGMTTTPVHSIFLGHFTFLVNHRQASLEISPIFLTAQTLL 319
 QY 301 MDLGQFLPCHTSSHOHDGMEAYVYKVDSCPEPQLRMKMNEBAEDYDDDLTDSMDVVRP 360
 DB 320 MDLGQFLPCHTSSHOHDGMEAYVYKVDSCPEPQLRMKMNEBAEDYDDDLTDSMDVVRP 379
 QY 361 DDNSPSPTQIRSVAKKPKTWVHYTAABEEDWDYAPLVLAPDDRSYKSOVLNNGPQRTG 420
 DB 380 DDNSPSPTQIRSVAKKPKTWVHYTAABEEDWDYAPLVLAPDDRSYKSOVLNNGPQRTG 439
 QY 421 RYKVKYRFMAYTDETFKTRAIQHSIGLGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 480
 DB 440 RYKVKYRFMAYTDETFKTRAIQHSIGLGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 499
 QY 481 TDVRELYSRRLPKGVKHLKDFDILPGETFKYKWTVTVEGPTKSDPRCLTRYSSFVNME 540
 DB 500 TDVRELYSRRLPKGVKHLKDFDILPGETFKYKWTVTVEGPTKSDPRCLTRYSSFVNME 559
 QY 541 RDLASGLIGPLIACYKESVDQRGNQIMSDKRVILFSVFDENRSMYLTENIQRFPLNPAG 600
 DB 560 RDLASGLIGPLIACYKESVDQRGNQIMSDKRVILFSVFDENRSMYLTENIQRFPLNPAG 619
 QY 601 VOLEDPFQASIMESINGVYVDSQLSVCLHEVAYWILSTIGATDFLSVFSSGYTKH 660
 DB 620 VOLEDPFQASIMESINGVYVDSQLSVCLHEVAYWILSTIGATDFLSVFSSGYTKH 679
 QY 661 KMVYEDTLTLFPFSETVFMSENENGLMGLGCHNSDENRGWNTALLKYSSCDKNTGDYVE 720
 DB 680 KMVYEDTLTLFPFSETVFMSENENGLMGLGCHNSDENRGWNTALLKYSSCDKNTGDYVE 739
 QY 721 DSYEDISAYLLSKNNAIEPRSP----- 742
 DB 740 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTRQKFNATTIPENDIEKTDPAFHRTPMP 799
 QY 743 ----- 742
 DB 800 KIQNVSSDLMLLRQSPTPHGLSLSLDQEAKEYTFSDPSPGAIDSNNSLSMETHFRPQ 859
 QY 743 ----- 742

Db 860 LHHSGDMVFTPESSGLQRLNEKLCTTAATLKKLDPKVSTSNLLISTIPSDNLAAGTDN 919
 QY 743 ----- 742
 Db 920 TSSLGPPMPVHYDSQDITLFGKSSPLTESGGLSLENDSKLLRESGLANSQSSW 979
 QY 743 ----- 742
 Db 980 GKNVSTESGRLFKGKGAGHPALLTKDNALFKVISILKTKNTSNNATNKRKTHIDGPSL 1039
 QY 743 ----- 742
 Db 1040 LIENSPVQNIILESOTEPKKVTPLIHDMMLDKNATPALRLNHSNKTSSKNMEMVQOK 1099
 QY 743 ----- 742
 Db 1100 KEGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKNSLNSGGQSPKQLVSLGPEKSVEG 1159
 QY 743 ----- 742
 Db 1160 QNFLSEKNKVVVGKEFTKDVGLKEMVFPSSRNLFNLNLDLHNNTHNQEKIQBIEIK 1219
 QY 743 ----- 742
 Db 1220 KETLIQENVVLPOIHTVTGKTFMKNLFLLSLSTRQNVESYDGAAYPVLODFRSLNDSNR 1279
 QY 743 ----- 742
 Db 1280 TKKHTAHFKKGBEENLEGNGQTKQIVKACTTRISNPTSQNFVTRQSKKALQKFL 1339
 QY 743 ----- 742
 Db 1340 PLEETELEKRIIVDDTSTQSKNKKHLPSTLQIDYNEKEGAIQTSPSLSDCLTRSHSI 1399
 QY 743 ----- 742
 Db 1400 PQANRSLPIAKVSPSPSIRPIYLRVLFQDNSSHLPAASRYKKSGVQESSHFLQAKK 1459
 QY 743 ----- 742
 Db 1460 NNLSLAILTEMTQDREVGLSATNSVYKKVENTVLPKPDLPKTSKVELLPKVHI 1519
 QY 743 ----- 742
 Db 1520 YQKDLFFTETNSGPHGLDLVEGSLQGTGEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579
 QY 743 ----- 742
 Db 1580 LLDPLANDNHYGTQIPKEBWKSEKSPKTAPEKKKDTILSLNACSNHAIANEGQNK 1639
 QY 743 -----SQPPVLKRHRIBITRTTLQSDQEEIDYDDTISVEMKKEDFDIY 786
 Db 1640 EIEVTWAKQORTERLCSQNPVPLKRHRIBITRTTLQSDQEEIDYDDTISVEMKKEDFDIY 1699
 QY 787 DEDENQSPSFQKTHRYTAAVERLWDYGMSSPHVLNRAGSGVPOPKVVFQEF 846
 Db 1700 DEDENQSPSFQKTHRYTAAVERLWDYGMSSPHVLNRAGSGVPOPKVVFQEF 1759
 QY 847 GSFTQPLRGELNEHGLGLGPYRAEVEDNIMVTPRQASRPYSFYSSLIISYEEORQCA 906
 Db 1760 GSFTQPLRGELNEHGLGLGPYRAEVEDNIMVTPRQASRPYSFYSSLIISYEEORQCA 1819
 QY 907 EPRKNFVKQNETKTYFWKVOHMAPTKDFCDCKAMAYFSDVLEKDVHSLGLIGLIVCHT 966
 Db 1820 EPRKNFVKQNETKTYFWKVOHMAPTKDFCDCKAMAYFSDVLEKDVHSLGLIGLIVCHT 1879
 QY 967 NTLNPAHQVQVTOEALFTTFIDETKSWYFTENMERNCRAPCNIQMEDPTKENVYRHA 1026
 Db 1880 NTLNPAHQVQVTOEALFTTFIDETKSWYFTENMERNCRAPCNIQMEDPTKENVYRHA 1939
 QY 1027 INGYINDTLFGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTFRKKEEYKALYNLYP 1086
 Db 1940 INGYINDTLFGLVMAQDQIRWYLLSMGNSNENIHSIFSGHVTFRKKEEYKALYNLYP 1999

QY 1087 GUPETVEMLPKAGIWRVECLTGEHLHAGMSTLFLVYSNKCQTPPLWASHIRDFQITAS 1146
 Db 2000 GUPETVEMLPKAGIWRVECLTGEHLHAGMSTLFLVYSNKCQTPPLWASHIRDFQITAS 2059
 QY 1147 GQYGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGAROKFSSLYISQ 1206
 Db 2060 GQYGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGAROKFSSLYISQ 2119
 QY 1207 FIIMYSLGCKQGTGRNSTGTLMVFFGNVDSGIGKHNIFNPPPIIARYIRLHPHYSIRS 1266
 Db 2120 FIIMYSLGCKQGTGRNSTGTLMVFFGNVDSGIGKHNIFNPPPIIARYIRLHPHYSIRS 2179
 QY 1267 TLRMELMGCDLNSCMLPCHESKASDAQITASSYFTNNPATWSPSKARLHLCGRNAR 1326
 Db 2180 TLRMELMGCDLNSCMLPCHESKASDAQITASSYFTNNPATWSPSKARLHLCGRNAR 2239
 QY 1327 PQVNPKEWLQVDFQKTMKVGTGTVQGVKSLLTSMYVKEFLISSSDQGHQWTLFFQNGKV 1386
 Db 2240 PQVNPKEWLQVDFQKTMKVGTGTVQGVKSLLTSMYVKEFLISSSDQGHQWTLFFQNGKV 2299
 QY 1387 KVFQGNQDSFTPVNSLDPPLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 1438
 Db 2300 KVFQGNQDSFTPVNSLDPPLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY 2351

RESULT 2

T42763

coagulation factor VIII precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

C:Accession: T42763

R:Lollar, P.

submitted to the EMBL Data Library, August 1996

A:Reference number: Z22269

A:Accession: T42763

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-2133 <LOL>

A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PID:AA806705.1

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; p

F:1-15/Domain: signal sequence #status predicted <SIG>

F:20-2133/Product: coagulation factor VIII #status predicted <MAT>

F:23-349/Domain: ferroxidase repeat homology <FOX1>

F:402-730/Domain: ferroxidase repeat homology <FOX2>

F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match

Best Local Similarity 81.0%; Score 6227; DB 2; Length 2133;

Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

QY 1 ATRRYLGAVELSWDYMOSD-IGELPVDARPPRPVPSFPFNTSVYKTLFVEFTVHLF 59
 Db 20 ATRRYLGAVELSWDYMOSD-IGELPVDARPPRPVPSFPFNTSVYKTLFVEFTVHLF 79
 QY 60 NIAXERPMPGLGPTICQAEVYDVTWITLKNWASHPSLHGVSYWASGAEVDDOTS 119
 Db 80 SVARPRPMPGLGPTICQAEVYDVTWITLKNWASHPSLHGVSYWASGAEVDDOTS 139
 QY 120 QREKEDDDVFPFGSHYVYQVQLKENGPMASDPLCLTYSYLSHVDLVKDNLSGLIGALV 179
 Db 140 QREKEDDDVFPFGSHYVYQVQLKENGPMASDPLCLTYSYLSHVDLVKDNLSGLIGALV 199
 QY 180 REGSLAKETQTLHKFILLPAVDFEGKSWHSETKNSLMQDRDAASARAWPMHVTNGYVN 239
 Db 200 REGSLAKETQTLHKFILLPAVDFEGKSWHSETKNSLMQDRDAASARAWPMHVTNGYVN 259
 QY 240 RSLPGLGCHKSVYMWVIGTTPPEVHSIFLEGHTFLVRHHRQASLSISITFLTAQTL 299
 Db 260 RSLPGLGCHKSVYMWVIGTTPPEVHSIFLEGHTFLVRHHRQASLSISITFLTAQTL 319
 QY 300 LMDLQGLFLLCHISSHQDGMAYVYKVDSCPEEPQLRMKNNEAEYDDDLTDSMDVVR 359

2113	Db	FFGNVSSGIGKHSFPPIIARIYLRHPHSSIRSTRMLMGCDLNSCSIPLGMSKVI	2172
1292	Qy	SDAQITASSYFTNNFATWSPSKARLHQGRSNAWRPQVNNPKEMQLVDFOKTKMKVGTGVT	1351
2173	Db	SDQITITASSYFTNNFATWSPSKARLHQGRSNAWRPQVNNPKEMQLVDFOKTKMKVGTGVT	2232
1352	Qy	QGVKSLTSMVYKVEPLTSSSDGCHOWTLFPQNGKVKVFGQNGDSFTPVVNSLDPPLATRY	1411
2233	Db	QGVKSLTSMVYKVEPLTSSSDGCHOWTLFPQNGKVKVFGQNGDSFTPVVNSLDPPLATRY	2292
1412	Qy	LRHPQSWHQIALRMVILGCEAQDLY	1438
2293	Db	LRHPQIWEHQIALREILGCEAQOY	2319

RESULT 4

KRHS
coagulation factor V precursor [validated] - human
N/Alternate names: coagulation labile factor; proaccelerin
C/Species: Homo sapiens (man)
C/Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000
C/Accession: A56172; A42344; A28028; A27498; A25897
R/Cripe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A/Title: Structure of the gene for human coagulation factor V.
A/Reference number: A42344; MUID:92232688; PMID:1567832
A/Accession: A56172
A/Molecule type: DNA
A/Residues: 1-2224 <CRI>
A/Cross-references: GB:J05368
A/Accession: A42344
A/Molecule type: DNA
A/Residues: 48-58; 79-89; 120-130; 191-201; 239-249; 313-323; 368-378; 428-437; 461-471; 533-542;
2070; 2111-2120; 2172-2181 <CR2>
R/Jenny, R.J.; Pittman, D.I.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A/Title: Complete cDNA and derived amino acid sequence of human factor V.
A/Reference number: A28028; MUID:87260886; PMID:3110773
A/Accession: A28028
A/Molecule type: mRNA
A/Residues: 1-857; 'R', 859-964, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>
A/Cross-references: GB:M16967
A/Note: parts of this sequence, including the amino end of the mature protein, were deter
R/Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A/Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A/Reference number: A27498; MUID:98107560; PMID:2827731
A/Accession: A27498
A/Molecule type: mRNA
A/Residues: 1-1284, 'I', 1286-1600 <KAN>
A/Cross-references: GB:M17785
A/Note: parts of this sequence were determined by protein sequencing
R/Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A/Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A/Reference number: A25897; MUID:86313665; PMID:3092220
A/Accession: A25897
A/Molecule type: mRNA
A/Residues: 1188-1215, 1315-2224 <KA2>
A/Cross-references: GB:M14335
A/Note: parts of this sequence were determined by protein sequencing
R/Keller, P.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A/Title: Thrombin-catalyzed activation of recombinant human factor V.
A/Reference number: A56139; MUID:95210278; PMID:7696276
A/Contents: annotation; thrombin cleavage sites
C/Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C/Genetics:
A/Gene: GDB:F5
A/Cross-references: GDB:I119896; OMIM:227400
A/Map position: lq23-lq23
A/Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65

Qy 531 RYSSFYNNMRLASGLIGPLLCYKSSVDORGNOIMSKDKENVILSFVEDNRNRYLWLTEN 590
Db 503 RPYSDVDIMRDIASGLIGLLICKSRLDRRGQRAADIEQQAVFAVFDENKSWYLEDN 562
Qy 591 IORLEPNAGVQLEDPPEFOASNIHMSINGYVFDLSL-QLSVCLEHVAVYVYILSIGAQTDFL 649
Db 563 INKCEPNDEKRDPPKYESNIMSTINGYVPEISITTLGFCDFDVTQVHFCSVGTQNEIL 622
Qy 650 SVFSGYTFKHMYVEDTLTLPPSGEYTFVMSMENPGLWILGCHNS ----- 695
Db 623 TIHETGHSFYKGRHEDTLTLPKRGESVTVTMDNVGTWMLTSMNSPSRKLKLRFDV 682
Qy 696 -----DFNR----- 700
Db 683 KCIPTDDSDSEIPEPPSTWATRKMDRLLEPEDEESDADYDQNRLLAALGIRSPFNS 742
Qy 701 -----GMTAL----- 705
Db 743 SLNQREERFNLTALALENGTEFVSSNTDIIVGSNYSSPSNISKFTVNNLABEPQAPSHQ 802
Qy 706 -----LKVSCDKNTGDIYEDSD----- 725
Db 803 ATTAGSPLRLHIGKNSVLNSSTAESSPYSEDPIEDPLQPDVTGIRLLSLGAGEFKSQEH 862
Qy 726 ----- 725
Db 863 AKHKGPVERDQAQKHFPSMMLLAHKVGRHLSDQTGSPGHRPWEDLPQDTSGPSMR 922
Qy 726 -----ISAYLLSKNNA----- 736
Db 923 FWKDPSSLALLKQSSSKILVGRMHLASEKGYEIIQDTEDTAVNNWLLISPOQWASRAW 982
Qy 737 ----- 736
Db 983 GESTFLANKPGQSGHPKPRVRHKSQVLRQDQKSLKKSQFLIKTRKKKXKHTHAP 1042
Qy 737 IPRSF-----SQ 744
Db 1043 LSPRTFHLRSEAYNTFSERLKHSLVLAHKSNETSLPTDLNQTLPSPMDPGHIASLPDHNQ 1102
Qy 745 N----- 745
Db 1103 NSSNDTQASCPPLGLYQTVPEEHYQTFPIQDPQDMHSTSPSHRSSPSELMLEYDRS 1162
Qy 746 -----PVLKXH- 752
Db 1163 HKSFTDISQMSPSSEHEVMQTVISPDLSQVTLSPELSQTNLSPLDLSHTTILSPELIQRLN 1222
Qy 753 ----- 752
Db 1223 SPALQMPISPDLSHTTILSPDLSHTTILSLDLSQTNLSPELSQTNLSPALQOMPLSPDLSH 1282
Qy 753 ----- 752
Db 1283 TTLSLDLSQTNLSPELSHMTLSPELSQTNLSPALQOMPIPSDLSHTTILSLDLSQTNLSPE 1342
Qy 753 -----QREITRTTLQSQEHI----- 768
Db 1343 LSQTNLSPALQOMPLSPDPSHTTILSLDLSQTNLSPELSQTNLSPLDLSHTTILSPDL 1402
Qy 769 ----- 768
Db 1403 TPDLDQMTLSRDLGETDLSNFGOMLSPLDLSQVTLSPDLSHTTILSPDLSQISPPPLDQ 1462
Qy 769 -----DYDDT 773
Db 1463 IFYPSESSQSLLOBFNESFPYDPLGQMPSPSPPTLNDTFLSKFNPLVIVLSKDGTDY 1522
Qy 774 ISVENKKE-----DFDIYDE-----DENQSP-----RSFOKTRHY 804
Db 1523 IELIPKEVQSSDDYAEIDYVYDDPYKTVRTWINSRDPDNTAAWYLSNNGNRNY 1582
Qy 805 FIAAVERLDWYGMSSSPHVLNR--AQGSVQ-----FKKVVFEFTDGSFTQPLVRGBLN 859

Db 1583 YIAABEISWY-----SEFVQRETDIEDSDDIPEDTTYKVKVFKYLDSTFTKRPGRXEY 1638
Qy 860 EHLGLGPGYTRAEVEDNIMVTFNQASRPYSFYSSLSIYE-----EORQQAEPKRN 911
Db 1639 EHLGILGPIIRAEVDVIVQVRFKNLASRPYSLHAHGLSYEKSSEKGTYYEDDSEWPFKEDN 1698
Qy 912 FVKEKETTYFKVKQHEMAPTKQBFCKAWAYSDDVLEKDVHSGLLGPLLVCVHTNTLNP 971
Db 1699 AVQNSSTTYVWHAATERSGPESPGSACRAWAYSAVNPEDKIDHGLGLPPLLICQGLIHK 1758
Qy 972 AKGRQVTVQBFALFFTTIFDETKSYFTENMERNCRAPCNIQMEDPTTFKENVYRHAINGYI 1031
Db 1759 DSNMPLDMREVLFWTFDEKSNYEEKSSNR-----LTSSEMKKSHEFHAINGM 1812
Qy 1032 MDTLPLGLMAQDQIRIRWYLLSMGSENHISHPGSHVFTVRKKEEYKALNLYLPGVFET 1091
Db 1813 Y-SLPGKMYEQEWWRLHLNIGSQDIHVVFHFGQTLLENKQKHQGLVWPLLPSSFKT 1871
Qy 1092 VEMLPKAGIWRVECLIGELHLAGMSTLELVYSNKCOTPLGMSAGSHIRDFOITASGOVQ 1151
Db 1872 LEMKASPGWMLLNTVEGENQRAGNQTFPLMDRDCRMPGLSTGIISDSQIKASEFLGY 1931
Qy 1152 WAPKLARIHYSGSINAWSTKE---PFS---WIKVDLLAPMIHIGIKTQGARQKPSLYIS 1205
Db 1932 WEPLARLANNCGSYNAMSVEKLAAPASKPWIQVDMQKEVIITGTGTQAGHYLKSCYTT 1991
Qy 1206 QFIIMYSLDGKWKQTYRGNSTGTLWFFGNVDSSGIKHNFNPILLIARYLHHTHSIR 1265
Db 1992 EFYVAYSSNQINMQIFKGNSTRNVMYFNGNSDASTIKENQDPPDIVARIYIRISPTAYNR 2051
Qy 1266 STLMEMLMGCDLNSCSPGLGWESKAISSAOITASSSYTNMPAT-WSPSKARHLHQGRNA 1324
Db 2052 PTLLELQGCENCCSTPLGMEKIKENQITASSFKKSWGWDYWEFRALNACQVRNA 2111
Qy 1325 WRPOVNNPKMLQVDFQTKVKTGVTTCQGVKSLTSMYKFLISSQDGHQWTLFFQNG 1384
Db 2112 WQAKANNKQWLEIDLKIKITAIITQGCCKSLSESMYKSYTHYSEQGVEMKPYRLKS 2171
Qy 1385 KV--KVFQGNQDSTFPVNSLDPPLPLRYLRIHPQSWHQIALRMEVLGCEAQDLY 1438
Db 2172 SWMDKIPEGNTNTRKGVKNFNFNPIISRFIRVPTKNTNQSLALRLEFGC---DIY 2224

RESULT 5

KFB05

coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999

C:Accession: A42580; A36497

R:Guinto, E.R.; Eamon, C.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A:Title: The complete cDNA sequence of bovine coagulation factor V.

A:Reference number: A42580; MUID:92147638; PMID:1737753

A:Accession: A42580

A:Molecule type: mRNA

A:Residues: 1-2211 <GUI>

A:Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038

A:Note: Sequence extracted from NCBI backbone (NCBIN:80774, NCBIP:80776)

R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A:Title: Identification and characterization of a phospholipid-binding site of bovine

A:Reference number: A36497; MUID:91072354; PMID:2254316

A:Accession: A36497

A:Molecule type: protein

A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X',

R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochemistry 33, 13109-13116, 1994

A:Title: Determination of the disulfide bridges in factor Va heavy chain.

A:Reference number: A55979; MUID:95034740; PMID:7947716

A:Contents: annotation

A:Note: 566-Cys and 617-Cys were shown to have free sulfhydryls

C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.

C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
 C:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
 F:1-28/Domain: signal sequence #status predicted <Sig>
 F:29-221/Product: coagulation factor V #status predicted <Mat>
 F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
 F:29-345/Domain: A1 <DA1>
 F:33-329/Domain: ferroxidase repeat homology <FO1>
 F:346-695/Domain: A2 <DA2>
 F:351-688/Domain: ferroxidase repeat homology <FO2>
 F:696-1564/Domain: B <DOB>
 F:1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
 F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
 F:1565-1892/Domain: A3 <DA3>
 F:1572-1892/Domain: ferroxidase repeat homology <FO3>
 F:1654-1752/Region: phospholipid binding #status predicted
 F:1893-2051/Domain: C1 <DC1>
 F:1893-2048/Domain: discoidin I amino-terminal homology <DNI>
 F:2052-2211/Domain: C2 <DC2>
 F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>
 F:167-193,248-329,499-525/Disulfide bonds: #status experimental
 F:225-239,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,1094,
 F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363-697,1537/Binding site: sulfate (tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:533-534/Cleavage site: Arg-Gly (protein C) #status predicted
 F:607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
 F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
 F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
 F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 30.8%; Score 2372; DB 1; Length 2211;
 Best Local Similarity 26.2%; Pred. No. 1,1e-149;
 Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

QY	3	RYYLCAVLSNDYQMSDGLGELPVDARPPRPVKSPFNTSVYKTLFVETVHLFNIA	62
DB	32	RFYVAAQIRWNYR-----PESTHL-----SSKPFETS--FKKIVRYEAY-FQKE	76
QY	63	KRPFPMMGLGPTQIAEVYDVTITLKNMASHPVSLHVGVSVMKASEAGYDDDSORE	122
DB	77	KQSRTSGLLPTLYAEVGDIMKVHFKNAKHLPSLHAQIKYKFESEASDHLPMW	136
QY	123	KEDDKVPQGSHTYVWQVLEKNGPMASDPLCTYSVLSHVDLVKOLNSGLIGALLVCRG	182
DB	137	KMDDAVAPQGYTYEYIISEHSGPTDDPPCLTHIYYSVNLVEDFNSGLIGLLICKKG	196
QY	183	SLAKEKTQTL--HKFTLLFAVDEGKSHSETKNSLMQDRDAASARAMPKHTVNGYVNR	240
DB	197	TLTEOCTQWFKQVLMFAVDESKSNQTS-----LMTVNGYVNG	240
QY	241	SIFPLGCHRKSVYHVHVGITGTPPEVHSIFLEGHTFLVNRHQASLEISPIITFLTAQTL	300
DB	241	TWPDITVCAHDHISWHLIGMSGPELFSIFENGQVLEQNHKISAITLVSATSTANMTV	300
QY	301	MDLQGFLLCHTSSHQHDMWAYVKVDCSEFPQLEMKMNEAEYDDDLTUSEMDVVRP	360
DB	301	SPEGRWITASLIPRHFGAQWQYIDIKNCAKTRNPKK-----LTRDQ-----	343
QY	361	DDNDSPTFIQIRSVAKKHPTKTVHVIYAAEEEDWDVAPLVLPDDRYSKSOVLNNGPQRTG	420
DB	344	-----RRHKIRWEYFIAAEVLDVAPITIPANMDKKYRSLHUNFSNRIG	388
QY	421	RYKKVRFWAYTDTFKTR--BAIQHESGILGPLYGEVGDTLIIIFKNQASRPYNIYPH	478
DB	389	KHYKKVYQXQDSDSTKLEDPSSBEGDILGFIIIRAOVRDLKIVFKNMASRSYIVPH	448
QY	479	GLT-----DVRPLYSRLLPKGVKHLKDFILPGLIFPKYKWTVTVEDGTKSDPRCL	529
DB	449	GVTFSYDNEVNSSSTSGSNTMIRAVR-----PGETYTYKWNILESDEPTENDAQCL	500
QY	530	TRYSSFYNMERDLASGLIGPLLIYKESVDQRGNQIMSKRNVILFVDFENRSWLYTE	589
DB	501	TEPYSNVDITRDLASGLIGLLICKSRSLDRGIQRAADIQQQAVPAVFDENKSWYIED	560
QY	590	NIQRLPNPAGVQLEDPFQASNIH-----SINGYVFDLSQ-LSVCLHEVAYWILSIG	643
DB	561	NIYKFCENPEKVKRDPDKFYFESNIMSNFTLPAINGVPEIPILGFCFDDTVQMHFCSVG	620
QY	644	AQTDPLSVFFSGYTFPKHWYVEDTLTLPPFSGETVFMENPGLMWILCHNSDFENRMT	703
DB	621	TQNDILTHFTGHSEFYGKRHEDTLTLFPMQGESVTVTMDNVGTWMTLTMNSNPKCLR	680
QY	704	ALLKVSCDKMTGD-----YVEDS-----YEDISAYLL--	731
DB	681	LRFRDAKIRDDDDSYBIIEYPSGSTAMTTKIHDSIEIEDNDADSDYDELALILGL	740
QY	732	-----SKNNAIEPRFSQ-----	744
DB	741	RSFRNSLNQSKDELNLTALALEKDSEFPFSANRSLDSNSSSRSHVSRLLAKNPAESLK	800
QY	745	-----NPPVLK-----	750
DB	801	TLLHLEAPAGSPLEHAGLDKNSALNPMAEHSPPSYSEDPREDHPLSDVTGVSLLPFCGTG	860
QY	751	-----RHOR-----EITRTTLOSQO-----EEI--	768
DB	861	FKNRKPAKHQRFQVGRQAQAKHKFSQTRFPAKHTRTRLSQDNSSSRMGPMEDIPSDILL	920
QY	769	-----	768
DB	921	LOQKDPYKILNGEHWLSEKSYELIQDANENKTVNKLPSNPQSDSRTPWAGENIPFKNSHG	980
QY	769	-----	768
DB	981	KQSGHTPLVTRERPLQDRDRNSRLKEGLPLIRTRKKKEKPAHVHPLSPRSFHLR	1040
QY	769	-----DYDD-----	772
DB	1041	GEVNASFDRDRHNSLLHASNETSLSIDLNQTFPSMNLASLAPDHDQTSNDTTSCQ	1100
QY	773	-----TISVENKKEPDFIYDEB-----ENQSP-----	794
DB	1101	SSPDLYTVSPBEHYQIFPIQDSDPHTSTAPNRSPPDHTSTAPNRSPPDHTSTAPNRSPP	1160
QY	795	-----	794
DB	1161	NYDLNRRAIPTDVQIIPSELEVWQATSLDLSQPSISPDLGQWALSPPDQESLSPLD	1220
QY	795	-----	794
DB	1221	QOTSPLDLSQESLSPLDQGTALSPDPQESLSPLDQGTALSPDPQESLSPLDQGTALS	1280
QY	795	-----	794
DB	1281	PDPQESLSPLDQGTSLSPDLSQESLSPLDQGTALSPDPQESLSPLDQGTALSPDPQES	1340
QY	795	-----	794
DB	1341	SLSPDLGQTSPLDQESLSPLDQGTALSPDPQESLSPLDQGTSLSPDLSQESLSPLD	1400
QY	795	-----	794
DB	1401	QGTALSPDLSQESLSPLDQGTPLSPDLSLSLSPLDQLDLKQTSPPDLNQTSHTESS	1460
QY	795	-----	794
DB	1461	QSLPLPRGQTFPNADICQMPSPDPSTLNTTFFBEFNPVWVGLSRDDGDIYEIIPRQ	1520
QY	795	-----RSFOKTRTHYFAIAVER	811
DB	1521	KESSSEEDYGEFFEVAYNDPYQTDLRDIDNSRNPNDIAAWYLRNSNTGNRKYYIAAEEI	1580
QY	812	LWDYGMSSPHVLNRASGS---VPQ---FKKVVFEFTDGSFTQPLYRGLNELHGLL	865

Db 1581 SMDYS-----KFGQSDVDVYVPEDTVYKVVFKYLDSTFTKLPDQGEYBEHLGIL 1631
Qy 866 GPYTRAEVEDIMWTFRQARSPISYSLISYE-----EDORQABPRKPFVKNPE 917
Db 1632 GPVIRAEVDVQVFRKPLASRPSYSLHAGLSYEKSSBKGYEDDSDPEWFKEDNAIQPNK 1691
Qy 918 TKTYPMKQHMMAPTKDFCDKAWAYFSDVDLEKDVHSGLGILGPLLVCHTNTLNPAHQVQ 977
Db 1692 TYTYVWHATTSGPENPQSCACAWAYYSAVPEKDIHSGLGILGPLLVCHTNTLNPAHQVQ 1751
Qy 978 TVQEPALFTTIDETKSYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTLPG 1037
Db 1752 DMREPVLLFMVDEKESYWKPKTRSWERASS-----EVKNSHEFAHNGYIMYIN-LPG 1804
Qy 1038 LVMAQDQRIKRWLLSMGNSNENHSHFSGHVFYTKKEEYKQALNLYPGVPEIVEMLPS 1097
Db 1805 LMYEQBWRLHLLNLGSGRDILHVHFGQTLLENGTQQHQLGWVPLLPFGSKTLEMKAS 1864
Qy 1098 KAGIWRVCLIGEHLHAGMSTLFLVYSNKCOTPLQMASGHIRDFOITASGQGVQWAPKLA 1157
Db 1865 KPGWLLDTEVEGIEIQAGWQTPFLIVDRECKWPMGLSTGLTADSQIQASEFWGYWEPKLA 1924
Qy 1158 RLHYSGSNAM-----STK-BPFSWIKVDLLAPMIHGIKQAGKQKSSLYSIFIMY 1211
Db 1925 RLNNGGYSNAMIAEKLTSEFPEPWIQVDMQKEVLLGTQAGAKHYLKPYTYTTEFCVAY 1984
Qy 1212 SLDGKKMOTYRGNSTGLWVFGNVDSSGIIKHNINENPPIARYIRLHPTHYSIRSLRME 1271
Db 1985 SLDRQWRIRFGNSTRNVMYFGNSDASTIKENQIDPPVAVKIRIISPTGSKYKPKALRLE 2044
Qy 1272 LMGDLNCSMPLGMESKAISDAQITASSYFTNMFAT-WSPSKARLHLQGRSNARWPOV 1330
Db 2045 LOGCEVNGCSTPLGMESKLENKQITASSFKSWGNWYEPFLAFLNAGQVNAQAKAN 2104
Qy 1331 NPKELQVDFQTKMTVGTQGVKSILTSMTYKVEFLISSQDGHQWTLFFQNGKV--KV 1388
Db 2105 NNNQWLQIDLLKIKITAITVQCKSLSEMTVSYTHYSQGTDMKPYREKSSMVDKI 2164
Qy 1389 FQGNODSTPPVNSLDPLLTLYLRIHPQSVHQALRWEVLGCEAQDLY 1438
Db 2165 FEGNNVGRGVKNFNPFIISIRKILIPKTNQSLALRLELFGC---DMY 2211
RESULT 6
T42764
coagulation factor V - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg, Blood 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by protein C
A:Reference number: Z22270, MUID:98282202, PMID:9616155
A:Accession: T42764
A>Status: preliminary; translated from G8/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:G3219690; PID:G3219691; PIDN:AAC99553.1
A:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F:350-682/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>
Query Match 30.8%; Score 2369; DB 2; Length 2183;
Best Local Similarity 25.8%; Pred. No. 1.7e-149;
Matches 577; Conservative 287; Mismatches 486; Indels 888; Gaps 31;
Qy 3 RRYLGAVELSDVMQSDLGELPVDARPPRPVPSFPFNSTVYVYKTLFVETFTVLEFNIA 62
Db 32 RQFYVAAQILWNHYPE-----PTDPSLNSIPSFKKIVIREVEQY-FKKE 75

Qy 63 KRPMPMGLLOPTIOAEYDVTWITLKNMASHPVSLHAGVSYWKASEGAEYDDOTSORE 122
Db 76 KPRSNGLLOPTIYAEVGVIVKHFRNKADKPLSIHQGIKYKSEKSEGASYADHTFPAB 135
Qy 123 KEDDKVPFGSGSHYTVQVLKENGPMASDPLCLTYSYLSHVDLVKDLNGLGALLVCREG 182
Db 136 RKDQAVAPGEYTYEWIVSEDSGPTDPPCLTHIYYSYENLTQDFNSGLGLPLICKKG 195
Qy 183 SLAKEKQTL--HKFILLFAVDEKSHSETKNSLMQDRDAASARAWPKMHTVGVNVR 240
Db 196 TLTEDGTQKQDFDKQHVLLFAVDESKS-----RSQPSLSMTYTINGFVNK 239
Qy 241 SLPLGLICHRKSVMVHVIGCTTPEVHSIFLEHTFLVRNHRQASLSPITFTTAQTLL 300
Db 240 TMDPLITCAHDHVSWHLIGMSSGPELPSIHENGQVLEQNKQKSVTVLVSATSTANWTM 299
Qy 301 MDLQOFLFLCHSHSHQHDGMEAYVKVDSCEPBPQLRMKNEEAEYDDDLTDSMDVVR 360
Db 300 SPEGRMIVSSLPKHYQAGMQAYIDIKNCPKTR----- 333
Qy 361 DDDNSPFIQIRSVAKKHPKTMVHYIAAEEDMDYAPLVLPDDRSYKSOVLNQPQIG 420
Db 334 ---SPK--TLTREQRIRMKWEYFIAABEVIMYAPVIPANMDKIYRSOHLDFNSQIG 387
Qy 421 RYKVKVRFMAYTDETF--KTRBAIQHESGILGPLLYGEVGTLLIIFKNOASRRPNIYPH 478
Db 388 KHYKVIYQVEETFTKTDNPSIKQSGILGPVIRAQVDTLKIIVKNASRYSIYPH 447
Qy 479 GITDVRFLYSRRLPKGV-----KHLKDPILPGEIIPKYKNTVTVVEDGPTKSPRLCTR 531
Db 448 GVT-----FSPYEDGINSSTSGSHITIRPVQGETTYKWNILEFDEPTENDAQCCLR 501
Qy 532 YSSFNWMDRLASGLGPLLI CYKESVDQRCNQIMSDKXNVILFSVDENRKSVLENI 591
Db 502 PYSDVDVTRDIASGLGLLICKSRSLDQGVORVADIEQAVFAVFDENKSYIEDNI 561
Qy 592 QRLPLNPAGVLEDPQASNMHSINGYVFDLSQ--LSVCLHEVAYVYILSIGAQTDFLS 650
Db 562 NKFCEPDVAKDDPKFYESNINSTINGYVPESTLGFCDPTVQHFVCSGVTHDILT 621
Qy 651 VFFSGYFKHKQVYEDTLTLPFSGETVFMSPENPGLWILGCHNSDFNRQMTALLKVSS 710
Db 622 IHTGHSFTYGRRHEDTLTFPMRGESVTVMADVGTWMLTMSNPKRRLRFRDVK 681
Qy 711 CDKNTGY-YEDSYE-----DISAYLIS 732
Db 682 CNR---DYDNEDSYEIEPPAPTMTTRIRHDSLENEFGIDNEDDDYQYLLASSLGIRSF 738
Qy 733 KNNAIKPR----- 740
Db 739 KNSSLNPEENFNLTALENSSEFISPTDVRVDSNSSRLSKIINNKLKDFORTLPGS 798
Qy 741 ----- 740
Db 799 GATVAGTLRLNLIGLDFNLNSSTHEHRSSTHEHNDMENFQSNITMVYLLPLGPKGSNR 858
Qy 741 ----- 740
Db 859 BQDKPKTIKYGRPHMKGRFSWKA PACKTGRHSNPNKNSYSGMKSEEDIPELIPLKQKI 918
Qy 741 ----- 740
Db 919 TSKFLNRRWRVASEKGSVEIIAANGEDTDVDKLTNSPQNQITVPRGESTSHNTTKPS 978
Qy 741 ----- 740
Db 979 DLPTFGVGHKSPHROEENSGFKRQLFTRTKKKKKKKLHLSPLSPRGDFPLRGN 1038
Qy 741 ----- 740
Db 1039 HSPFPDRRLNHSLLLHKSNETALSPLDNQTSFSMSTDRSLPDYNQYKNDTEQMSSLD 1098

```
QY 741 -----SFSQNPVVLKR----- 751
DQ 1099 LYQSVPAEHSPTFAQDPQTHSTTDPYRSPPELSQGLDYLSDHDFYDDIGLTSFF 1158
QY 752 ----- 751
DQ 1159 PDSQKSPSSDDQAIIPSSDLSLFTTISPDLQTIYPDLQDLQLLSPDNQKTSPPDLGQ 1218
QY 752 ----- 751
DQ 1219 VPLSPDDNQKTSPPDLGQVLSLSPDDNQKTSPPDLGQVLSLSPDDNQKTSPPDLGQVLSLSPD 1278
QY 752 -----HQR-----EIRFTTLOSQOE----- 767
DQ 1279 DQOMITSPDLGQVPLSSDNQKTSPPDLGQVPLFPEDNQYFYLDSQVPLSSDNQKTSST 1338
QY 768 ----- 767
DQ 1339 DLLTSLSPDQVLSLSPDQLPLPDSNQVTVSPDLSSLTLSPDFNEIILAPDLGQVLSL 1398
QY 768 ----- 767
DQ 1399 PDLIQTNPALNHGKASSADPDQASYPDSQASSLPDLNRLTHPHPPSPPTL 1458
QY 768 -----LDYDDT-----ISVEMKDEDDIYDEDENQSPRSFQ 798
DQ 1459 NNTLSRKFNPLVVLGSLRVGDGVVEIVSESPERIDEDYAEDDFVYNDPVRTDTRDV 1518
QY 799 KTRH-----YFTAAVERLWDYGMSSSPHVLNRRAQSGVPQ-----PK 837
DQ 1519 NSSRNPDTTAAWYLRGCHGKFFYIAEETIYNABPAQSEM--DHEDTGTPTKDTYK 1576
QY 838 KVFQBTGSGFTQPIYRGLNEHLGLGPIYRAEVEDNMVTFRQAQRPYFYSLSLIS 897
DQ 1577 KVFYKYLSTSTSRDPRAEYEBHLGILGPIYRAEVEDDVIQVRFKGLASRPYSLAHGUIS 1636
QY 898 YE-----EDQROGAEPKQNPVKNETKTFPWKQHEMAPKQDFCKANAYFSDVLD 949
DQ 1637 YEKSSGKYVEDSPWFQDDAVQNSVTVVHATKSGPNSPOSACRWAYASVNV 1696
QY 950 EKDVHSLGLPLAVCHTNTLNPAGHQVTVQEPALFTTIDETKSYFTFENBERNCRAPC 1009
DQ 1697 ERDIHSLGLPLLCRKGTLHMRNPLPMDREPVLLFMVFEKKSUYEKS--KGSR--- 1751
QY 1010 NIQMEDPTKYNRFAINGYIMDTLPGVMAQDQIRNYLLSMGSENIHHSCHVFP 1069
DQ 1752 --RIESPEKNAHKFYAINGMIYN--LPGLRMVEQEWVRLHLLNMGSGSDIHHVHFGQTL 1808
QY 1070 TVRKKEEYKMALYNLYPGVFETVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYSNKQOT 1129
DQ 1809 LDNRKQHQGLGVWFLPGLSPKTLKEMKSKPGWLLDTEVGENQVAGMQTFFLLIDKECKM 1868
QY 1130 PLGNAGSHRIDPOITAGQGVQWAPKLARLHYSGSINAWSTKE-----PFS-WIKVDLLA 1183
DQ 1869 PMGLSTGVISDSQIKASEYLLTWEPRLARLNAGSYNAWSIEKTLDPFKIOWDMQK 1928
QY 1184 PMIHGKTCGAPQKSSYVISOFTIMYSIDGKQWYRGNSGTGLWVFFGNDVSSGKHX 1243
DQ 1929 EVVVTGICQGAHYLKSCTTTEFQVAYSDDQTNMQIFRKSGKSNVYFTGNSDGTIKE 1988
QY 1244 NIFNPPIIARYILRHPHTHYSIRSLRMLMGLCDLNSCPLMGESKAISDAQITASSYFT 1303
DQ 1989 NRLDPPVARYIRIHTKSYNRPTRLALELQGCVEVNGCSTPLGLEDGRIQDKQITASSPK 2048
QY 1304 NMPAT--WSPSKARLHLQGSNAWRPQNNPKHLOVDFOKTMKVQVTTQGVKSLTSMY 1362
DQ 2049 SWMGDYWEPSLARLNAQGRVNAWQAKANNKQVLDLAKIKKVTVAIVTQGCCKSLSEMY 2108
QY 1363 VKFEFLSSSDQGHQWTLFFQNGKV--KVFGQNDSDFTPVVNSLDPLLTATYLRHHPQSVW 1420
DQ 2109 VKSYSTQYSDQGVAMKPYRQKSMVDKIFEGNSNTGKMKNFNPIISFRIRIIPKTNW 2168
QY 1421 HQIALRMEVLGCEAQDLY 1438
```

```
DB 2169 QSIALLRLFLFC---DIY 2183
```

RESULT 7

KUUU

ferroxidase (EC 1.16.3.1) precursor [validated] - human
N:Alternate names: ceruloplasmin
N:Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)
C:Species: Homo sapiens (man)
C:Date: 31-Aug-1980 #sequence revision 12-May-1995 #text_change 08-Dec-2000
C:Accession: A25443; A24165; A35450; A00524; I59067
R:Kochinsky, M.L.; Funk, W.D.; van Oost, B.A.; Macgillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986
A:Title: Complete cDNA sequence of human ceruloplasmin.
A:Reference number: A25443; MUID:86259737; PMID:2873574
A:Accession: A25443
A:Molecule type: mRNA
A:Residues: 1-1060,1065-1069 <KOS>
A:Cross-references: GB:M13699; NID:g180255; PIDN:AAA51976.1; PID:g180256
A:Note: this is the short or CP-2 alternatively spliced form
R:Merz, J.F.B.; Grimes, A.
FEBS Lett. 203, 185-190, 1986
A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal le-
A:Reference number: A24165; MUID:86275241; PMID:3755405
A:Accession: A24165
A:Molecule type: mRNA
A:Residues: 1-40;549-599;784-829;919-952 <MER>
R:Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.
J. Biol. Chem. 265, 10780-10785, 1990
A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by alt
A:Reference number: A35450; MUID:90285218; PMID:2355023
A:Accession: A35450
A:Molecule type: DNA
A:Residues: 1007-1064 <YAN>
A:Cross-references: GB:J05506
A:Note: this is the long or CP-1 alternatively spliced form
R:Takahashi, N.; Otzel, T.L.; Putnam, F.W. 1984
Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984
A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequen-
A:Reference number: A00524; MUID:84119493; PMID:6582496
A:Accession: A00524
A:Molecule type: protein
A:Residues: 20-1060,1065-1069 <TAK>
A:Note: 79-Gly and 449-Gly were also found
R:Yang, F.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McGi-
Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A:Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A:Reference number: I59067; MUID:86205876; PMID:3486416
A:Accession: I59067
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 218-1069 <RBS>
A:Cross-references: GB:M13536; NID:g180248; PIDN:AAA51975.1; PID:g180249
C:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions p
C:Comment: In Wilson's disease the plasma levels of ferroxidase are diminished or unde
C:Comment: The three fragment chains are produced spontaneously during purification an
C:Genetics:
A:Gene: GDB:CP
A:Cross-references: GDB:I19069; OMIM:I17700
A:Map position: 3q23-3q25
A:Introns: 1006/3; 1061/1
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the
A:Note: iron(III), but not iron(II), is the form bound and transported by transferrin
A:Note: other possible functions are amine oxidase activity, copper transport and home
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxid
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1069/Product: ferroxidase, long form #status predicted <MATS>
F:20-1060/Product: ferroxidase, short form #status experimental <MATS>
F:20-499/Product: ferroxidase 67K chain #status experimental <K67>

```

Query Match          22.5%; Score 1731.5; DB 1; Length 1069;
Best Local Similarity 32.8%; Pred. No. 2.4e-107;
Matches 385; Conservative 195; Mismatches 409; Indels 193; Gaps 22;

QY      3 RRYILGAVELSDWYMSQSLDGE---LPPDARFPFPRVPKSPFFNTSVVVKTKLFVFBEFVHLF 59
DB      22 KHYIIGLIETWDY-ASDHGEKKLIISVDTEHSNIYLONGPDRIGRLYKKALYYOYTDEIF 80
QY      60 NTAKPFPWMGLLPPTIOAEIVYTVITLKMAASHPVSLHAGCVSTWKASEGABEYDDQTS 119
DB      81 RTTIEEPVMGLFGPTIKAEATGDKVYVHLKNLASRFTFTHSHGITTYIKHEGAIFYPDNNT 140
QY      120 QREKEDKYFPGSGSHYTVWQVLKENGFMASDPCLCTVSYLSHVLDVKOLNSGLIGALLVC 179
DB      141 DQRADDKVYPGEQYTYMMLATEEQSPGEDGCNVTRIYHSHIDAPKD TASGLIGPLIIC 200
QY      180 REGSLAKETQTL-HKPIILLFAVDFEGSKWHSE-----TKNSLMQDRDAASARA WPKMH 232
DB      201 KKDSLCKEKKHIDREFVVMFVSVDENFNFWLESDNIKTYCSEPEKYDKDNEDFQESNRMY 260
QY      233 TVNGYVNRSLPGLIGCHRSYVHWIYGMGTTPVAHSIFLEGHTFLVNRHQASLESPTT 292
DB      261 SVNGYTFGSLFGLSMCAEDRVKRWYLFMGMEVDVHAAPFHGOALTWNKTTRIDTINLFPAT 320
QY      293 FLTAQTLLMDLGOLFLLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNQNEAEYDDDLTD 352
DB      321 LPDAYVVAQNPGEWMLSCQNHLKAGLAQAFQVQEC-----NKSSKD----- 364
QY      353 SEMDVVRFPDDNSPFIQTRSVAKKHPKWVHVIIAAEEEDNDYAPL-----VLADP 403
DB      365 -----NIRKGHVRH-----YYIAAEEITWNYAPSGIDIFTKENLTAFC 402
QY      404 DRYSKSOYLNNGPQRIGRYKVKRFMAYTDEF-----KTREAIQHESGILGPLLYGVWGDT 460
DB      403 SDS--AVFPEQGTRIGGSYKLVAREYTDASTFKNERKERPBEEHGLIGLPVIWAEGVT 460
QY      461 LLIIFNQAASRPVNIYPHGI-----TDVRPLY---SRRLPKGVKHLKDPFILGEIF 509
DB      461 IRVTFHNKGAYPLSIPBIGVFRFNKNNEGTYISENTYNPQRSVPSPASH-----VAPTETF 515
QY      510 KYKWTTVSDGGPKSPRCILTRYYSFVNAMERDLASGLIGPLLICYKESVDQGNQIMSD 569
DB      516 TWENTYVKEVGPNTADPCLAKYIYSAVDPDKDIFTLGPMKICKKSGSLHANGROKVD 575
QY      570 KRNVILFSVDENRSWYLTENIORFLPNFAGVOLEDPEFQASNIMESHINGYFDSLQ-LS 628
DB      576 KEFLPFTVFDENESLLEDNIRMFTAPDOVDKDEDQFESNMKMSHMGMFYGNQPGLT 635
QY      629 VCLHEVAYWILSIGAQTDPLSVFRFGYFKHKMYVEDILTLPFPFSGETVFMSEMNGULM 688
DB      636 MCKGDSVVWYLFSGAGNEADVHGIFYSGNTYLMRGERRDTANLFPQTSLSLHMWPDEGTF 695
QY      689 ILGCHNSDFENRGWALTALKVSSCDKMTGDYVEDYEDISAYLLSKNAIEPRSFSONPPV 748
DB      696 NVBCLTIDHYTGMMKQKYTVNQCRQS-----EDS----- 725
QY      749 LKRHQREIRFTTLQSQOEIEDYDDTIISVEKKKEDFDYDEDENQSPRSFQKTRHYFIAA 808
DB      726 -----TWIGERTYVIAA-----TWIGERTYVIAA 738

```

Query Match 21.9%; Score 1687.5; DB 1; Length 1059;
Best Local Similarity 32.5%; Pred. No. 2.1e-104;
Matches 381; Conservative 186; Mismatches 419; Indels 185; Gaps 22;

QY 3 RRYVLGAVELSWYMQ-SDLGEL-PVDARPPRPVKSPFNTSVVTKTLFVETVHLFN 60
DB 22 KHYIIGITEAVNDYASGESEKELISVDTEQSNFYLRNGPDRIGRKVKKALYSYTDGTF 81
QY 61 IAKPRPPWMLGLGFTTQABVYDVTWTLKNMASHPVSLHAGVSYWKASGABYDDQTSQ 120
DB 82 KTDKKNALGLFGEVIAKAVGDKSVHVKVFAFPYTFHAGVYTKANEGALYDPTD 141
QY 121 REKEDKVPFGSGSHYVWQVKNMGASDPLCLTYSLYSHVLDVKNLNSGLGALLVCR 180
DB 142 FORADDKLFPGQOQYLYLRA-NEFSPGEGDGNVTRHYSHVDAPKDIALSLGLPLTLCK 200
QY 181 EGSIAKEKQTL-KHFTLLFAVDEGKSHSEKNSLM-----QSDAASARAWPMHT 233
DB 201 KGSLLHKEKENIQEFVLMFVVDENLSWLENIKTFCEPEKVDKDNEDFQESNRMS 260
QY 234 VNGVNRSLFGLIGCHRSYVMHVGIMGTTPVHSIFLEGTFLVNRHROASLEISPTF 293
DB 261 INGVTGSLFGLSCARDRVKWYLCMGNEVDVHSALFHCAULTSKNYHTDIINLFPATL 320
QY 294 LTAQTLMDLQGLFLFCHISSHQHCHAVYKVDSCPEEPQLRMKNBERADVDDLTDS 353
DB 321 IDVSMVAQNPGVWMLSCQNLHLKAGLQAFQVRDC-----NKPSD-DDDIQDR 368
QY 354 EMDVVRDDNSPFIQIRSVAKKHPTWVHYIAAEEEDWDYAP-----LVLPAP 402
DB 369 HV-----NH-----YYIAAETIMDYAPSGTDTTGNLSLGS 402
QY 403 DRSYSQYLNNQPGRIKRYKVRWYATDETF---KTRAIQHSGSGLPLLYGEVGD 459
DB 403 DSRVFEQ-----GATRIGGSYKLVYREYTDSTFNKRKRGPDDEHLGILGPVIAEVD 458
QY 460 TLLIIFKQASRPNIYPHGITVR-----PLYERRLPKGVKHLKDPILPGLPKTKWTV 515
DB 459 IIRVTFNKQFPLSLQPMGVRETKENEGTYG---PDGRSSQKASHVAPKETFTYEWT 515
QY 516 TVEDGPTKSPCLTRYSSFVNMERDLASGLIGPLLCYKESVDQKGNQIMSDKRNVL 575
DB 516 PKWGPYADPVLCKMYISGVDLTKDIFGLGPMKICKSGLLADGRQKVDKPFYLF 575
QY 576 FSVFDENRSLYNTENIORFENPAGVQLEDPEFOANIMHSINGYVFDLSQ-LSVCLHEV 634
DB 576 ATVFDENESLLDNIAMFTAPENVNKEDEDFQESNMKSMNGMYGNLPLGNMCLGES 635
QY 635 AYWYLSIGAQTDFLSVFFSGSYFKHKWYEDTLTPFPFSGETVEMSMENPGLWILGCH 694
DB 636 IVWYLSAGNEADVHGIYFSGNTYLSKGERDANLFPKSLTLLMTPDTEGSDVECLT 695
QY 695 SDFNRGWTALLKVSCKDNTGDIYEDSYEDISAYLLSKNNAIEPRSFQNPVLRKHQR 754
DB 696 TDHYTGGMKQYTVNQ-----KQPEDVTLY----- 722
QY 755 EITRTTQSDQEBIDDDTISVENKEDDFDIYEDENQSPRSQKTRHYFIAAVERLMD 814
DB 723 -----QERTYVIAAVERVMD 738
QY 815 YGMSSSPHVLNRQAQSGV-----PQKVKVFEPTDGSFTQPLVYRGELNEH 861
DB 739 YSPSRDWEMLHLQEQNVSNAPLKDKEEFFIGSKYKVVREYTDSTFREQVKRAEBEH 798
QY 862 LGLLGPVIRAEVDNIIMVTFNRQAQSRPYSYSSLSIYEBDQKQABPRKNFVK---PNET 918
DB 799 LGILGPLIHADVAKVAVPKMATRPYSIHA-----HGVTKSTVAPTAPGEV 848
QY 919 KTFVFKVQHMATKDEDFCKANAYSDVLEKDVHSGSLIGPLLVLC---HTNLTNPAHR 975
DB 849 RTTYVQIPERSGAGTDSQIPWAYYSYVDRVDRKDLISGLIGPLIVCKSYKVKVFNPK--- 905
QY 976 QVTVQEPALFTTIFDETKSYFTTENMERNCRAPCNQIMEDPTFKENVRFAHNGYIMDTL 1035

DB 906 --KMEFSLFLVFDENBSWYLDNDINTYDPPEKVNKNDBEFIESNKHAINGKMGFNL 963
QY 1036 PGLVNAQDQRIWYLLSMGSENINISHPSCHVFTVRKKEEYKMALYNLYPGVFEVEML 1095
DB 964 QGLTHVHGDEVWYVAMGNEIDLHTVHPHGSFOYKXRGHSSVDVDFPFGTYQLEMP 1023
QY 1096 PSKAGIWRVECLIGBHLHAGMSTFLVYSNK 1126
DB 1024 PTPGTWLLCHVTDHAGMWTTVTLVFNQ 1054

RESULT 9
A44258
factor VII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258
R:Levinson, B.; Kenrick, S.; Gamel, P.; Fisher, K.; Gitschier, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386; PMID:1427887
A:Accession: A44258
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
C:Cross-references: GB:M90707; NID:g182316; PID:AAA58466.1; PID:g182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology <DN2>

Query Match 14.4%; Score 1104; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.8e-66;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 VFGNVDSSGIGKHNFNPPIIARYIRLPHTHYSIRSLTLMELMGCDLNSCSMPLGMSKA 1290
DB 9 VFGNVDSSGIGKHNFNPPIIARYIRLPHTHYSIRSLTLMELMGCDLNSCSMPLGMSKA 68
QY 1291 ISDAQITASSYFTNNPATWSPSKARLHLQGRSNMRPQNNPKWLQVDFQKTMKVTGYT 1350
DB 69 ISDAQITASSYFTNNPATWSPSKARLHLQGRSNMRPQNNPKWLQVDFQKTMKVTGYT 128
QY 1351 TQGVKSLTSMYKVEFLISSQDGHQWTLFFQNGKVKVFGNQDSFTPVNSLDPLPLTR 1410
DB 129 TQGVKSLTSMYKVEFLISSQDGHQWTLFFQNGKVKVFGNQDSFTPVNSLDPLPLTR 188
QY 1411 YLRHPQSWVHQIALRMEVLGCEADLY 1438
DB 189 YLRHPQSWVHQIALRMEVLGCEADLY 216

RESULT 10
JC4915
aga protein precursor - rat
N:Alternate names: O-acetyl-Ga3 ganglioside
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
R:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; MUID:96374422; PMID:8780713
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DDBJ:D84068; NID:g1620006; PIDN:BAA12210.1; PID:g1620007
A:Experimental source: CST cell
C:Comment: This protein is required for the O-acetylation of distal ganglioside stialic
C:Genetics:
A:Gene: aga
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homo-

F:1-21/Domain: signal sequence #status predicted <SIG>

F:28-60/Domain: EGF homology <EG1>

F:68-107/Domain: EGF homology <EG2>

F:110-267/Domain: discoidin I amino-terminal homology <DN1>

F:271-427/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.6%; Score 663; DB 2; Length 427;

Best Local Similarity 40.3%; Pred. No. 1.6e-36;

Matches 142; Conservative 61; Mismatches 12; Indels 26; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 78 KCLVTEDTQSG--DFTYIYCQPVGYSGIHCELGCSYKLGEGAIADSIASVYWG 135

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTOGARQKFSLY 1203

DB 136 FMGLQWGPFLARLYRTGIVNAWTASSYDSKFPQVDFLRKVRVSVMTQASRAGRAY 195

QY 1204 ISQFIIMYSLDGKKWQYRGSTGLMVFPGVDSGGIKNIENPPIIARVIRLPHYTHS 1263

DB 196 LKTFKVAISLDGRRFEFTQDESGTQKDFMGQDNNSUKINFNPTLRAQYIRLIPVBSCH 255

QY 1264 IRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFT--NMFA--TWSPSKARLHLQ 1320

DB 256 RGCTLRFELLCGELGSGEPLGKNTIPDSQITASSYKTNLRAFQWPHLGLDNGQ 315

QY 1321 RSNARWPQNNPKWQLVDQFQTKMVTGTTQGVKSLTSTMYKSEFLISSQDGHQWTLF 1380

DB 316 KINAWTAQNSAKWQLVDLQTKRVTGIITQAGDFGHIQYVASYKVAHSDGQVQWTVY 375

QY 1381 FQNGKVKYFQGNQDSFTPVNSLDPLRLRYLRIHPQSVWVHOIALRMEVLGC 1432

DB 376 BEQGTSTKVFQGNLDNNSHKKNIFERFPFARYVVRVPLSWHNRITLRLLELGC 427

Query Match 8.5%; Score 657; DB 1; Length 463;

Best Local Similarity 42.2%; Pred. No. 4.7e-36;

Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1124 SNKCOTPLGWSGHIRDFOITASGOY-----GQWAPKLARLHYSGINAW--STKEPFS 1175

DB 145 ASRCSTQLGWSGAIADSIASVYWGFMGLQWGPFLARLYRTGIVNAWHASNYDSKP 204

QY 1176 WIKVDLLAPMIHGIKTOGARQKFSLYISQFIIMYSLDGKKWQYRGSTGLMVFPGN 1235

DB 205 WIQVLLRKRVRVSVMTQASRAGRAYLTKFVAYSLDGRKFETQDBSGD--KEFLGN 263

QY 1236 VDSGSGKNIENPPIIARVIRLPHYTHSIRSTRLMELMGCDLNSCSMPLGMSKASDAQ 1295

DB 264 LONNSLUKNVNFPTLRAQYIRLIPVBSCHRGCTLRFELLCGELGCEPLGKNTIPDSQ 323

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYICKPHGYTGIIHCEIICNAPLGMETGAIDFQISASSMHLG 115

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTOGARQKFSLY 1203

DB 116 FMGLQWGPFLARLYRTGIVNAWTASSYDSKFPQVDFLRKVRVSVMTQASRAGRAY 175

QY 1204 ISQFIIMYSLDGKKWQYRGSTGLMVFPGVDSGGIKNIENPPIIARVIRLPHYTHS 1263

DB 176 MKTFKVAISLDGRRFEFTQDESGTQKDFMGQDNNSUKINFNPTLRAQYIRLIPVBSCH 235

QY 1264 IRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTN--MFATWSPSKARLHLQ 1320

DB 236 RGCTLRFELLCGELGSGEPLGKNTIPDSQITASSYKTNLRAFQWPHLGLDNGQ 295

QY 1321 RSNARWPQNNPKWQLVDQFQTKMVTGTTQGVKSLTSTMYKSEFLISSQDGHQWTLF 1380

DB 296 KFNWTAQNSAKWQLVDLQTKRVTGIITQAGDFGHIQYVASYKVAHSDGQVQWTVY 355

QY 1391 FQNGKVKYFQGNQDSFTPVNSLDPLRLRYLRIHPQSVWVHOIALRMEVLGC 1432

DB 356 RDOGALEGIKIPFGLDNNNSHKKNMFETPFLRFRVRLPVAWHNRITLRLLELGC 409

Query Match 8.5%; Score 657; DB 1; Length 463;

Best Local Similarity 42.2%; Pred. No. 4.7e-36;

Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1124 SNKCOTPLGWSGHIRDFOITASGOY-----GQWAPKLARLHYSGINAW--STKEPFS 1175

DB 145 ASRCSTQLGWSGAIADSIASVYWGFMGLQWGPFLARLYRTGIVNAWHASNYDSKP 204

QY 1176 WIKVDLLAPMIHGIKTOGARQKFSLYISQFIIMYSLDGKKWQYRGSTGLMVFPGN 1235

DB 205 WIQVLLRKRVRVSVMTQASRAGRAYLTKFVAYSLDGRKFETQDBSGD--KEFLGN 263

QY 1236 VDSGSGKNIENPPIIARVIRLPHYTHSIRSTRLMELMGCDLNSCSMPLGMSKASDAQ 1295

DB 264 LONNSLUKNVNFPTLRAQYIRLIPVBSCHRGCTLRFELLCGELGCEPLGKNTIPDSQ 323

QY 1296 ITASSYFT--NMFA--TWSPSKARLHLQGRSNAPVQNNPKWQLVDQFQTKMKTGVTQ 1352

DB 324 MSASSSYKTWNLRAPFGWPHLGRDNGCKINAWTAQNSAKWQLVDLGTQRTQVTG 383

QY 1353 GVKSLTASMYKSEFLISSQDGHQWTLFFQNGKVKYFQGNQDSFTPVNSLDPLRLRYL 1412

DB 384 GARDFGHIQVYESYKVAHSDGQVQWTVYEBQSGSKVFGNLDNNSHKKNIFEKDFMAYV 443

QY 1413 RIHPQSVWVHOIALRMEVLGC 1432

DB 444 RLVFVSWHNRITLRLLELGC 463

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYICKPHGYTGIIHCEIICNAPLGMETGAIDFQISASSMHLG 115

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTOGARQKFSLY 1203

DB 116 FMGLQWGPFLARLYRTGIVNAWTASSYDSKFPQVDFLRKVRVSVMTQASRAGRAY 175

QY 1204 ISQFIIMYSLDGKKWQYRGSTGLMVFPGVDSGGIKNIENPPIIARVIRLPHYTHS 1263

DB 176 MKTFKVAISLDGRRFEFTQDESGTQKDFMGQDNNSUKINFNPTLRAQYIRLIPVBSCH 235

QY 1264 IRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTN--MFATWSPSKARLHLQ 1320

DB 236 RGCTLRFELLCGELGSGEPLGKNTIPDSQITASSYKTNLRAFQWPHLGLDNGQ 295

QY 1321 RSNARWPQNNPKWQLVDQFQTKMVTGTTQGVKSLTSTMYKSEFLISSQDGHQWTLF 1380

DB 296 KFNWTAQNSAKWQLVDLQTKRVTGIITQAGDFGHIQYVASYKVAHSDGQVQWTVY 355

QY 1391 FQNGKVKYFQGNQDSFTPVNSLDPLRLRYLRIHPQSVWVHOIALRMEVLGC 1432

DB 356 RDOGALEGIKIPFGLDNNNSHKKNMFETPFLRFRVRLPVAWHNRITLRLLELGC 409

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYICKPHGYTGIIHCEIICNAPLGMETGAIDFQISASSMHLG 115

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTOGARQKFSLY 1203

DB 116 FMGLQWGPFLARLYRTGIVNAWTASSYDSKFPQVDFLRKVRVSVMTQASRAGRAY 175

QY 1204 ISQFIIMYSLDGKKWQYRGSTGLMVFPGVDSGGIKNIENPPIIARVIRLPHYTHS 1263

DB 176 MKTFKVAISLDGRRFEFTQDESGTQKDFMGQDNNSUKINFNPTLRAQYIRLIPVBSCH 235

QY 1264 IRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTN--MFATWSPSKARLHLQ 1320

DB 236 RGCTLRFELLCGELGSGEPLGKNTIPDSQITASSYKTNLRAFQWPHLGLDNGQ 295

QY 1321 RSNARWPQNNPKWQLVDQFQTKMVTGTTQGVKSLTSTMYKSEFLISSQDGHQWTLF 1380

DB 296 KFNWTAQNSAKWQLVDLQTKRVTGIITQAGDFGHIQYVASYKVAHSDGQVQWTVY 355

QY 1391 FQNGKVKYFQGNQDSFTPVNSLDPLRLRYLRIHPQSVWVHOIALRMEVLGC 1432

DB 356 RDOGALEGIKIPFGLDNNNSHKKNMFETPFLRFRVRLPVAWHNRITLRLLELGC 409

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYICKPHGYTGIIHCEIICNAPLGMETGAIDFQISASSMHLG 115

QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTOGARQKFSLY 1203

DB 116 FMGLQWGPFLARLYRTGIVNAWTASSYDSKFPQVDFLRKVRVSVMTQASRAGRAY 175

QY 1204 ISQFIIMYSLDGKKWQYRGSTGLMVFPGVDSGGIKNIENPPIIARVIRLPHYTHS 1263

DB 176 MKTFKVAISLDGRRFEFTQDESGTQKDFMGQDNNSUKINFNPTLRAQYIRLIPVBSCH 235

QY 1264 IRSTRLMELMGCDLNSCSMPLGMSKASDAQITASSYFTN--MFATWSPSKARLHLQ 1320

DB 236 RGCTLRFELLCGELGSGEPLGKNTIPDSQITASSYKTNLRAFQWPHLGLDNGQ 295

QY 1321 RSNARWPQNNPKWQLVDQFQTKMVTGTTQGVKSLTSTMYKSEFLISSQDGHQWTLF 1380

DB 296 KFNWTAQNSAKWQLVDLQTKRVTGIITQAGDFGHIQYVASYKVAHSDGQVQWTVY 355

QY 1391 FQNGKVKYFQGNQDSFTPVNSLDPLRLRYLRIHPQSVWVHOIALRMEVLGC 1432

DB 356 RDOGALEGIKIPFGLDNNNSHKKNMFETPFLRFRVRLPVAWHNRITLRLLELGC 409

Query Match 8.5%; Score 650; DB 2; Length 409;

Best Local Similarity 39.3%; Pred. No. 1.1e-35;

Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGELHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149

DB 58 ECEVIDDAHRG--DVFTYICKPHGYTGIIHCEIICNAPLGMETGAIDFQISASSMHLG 115

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <AOK>

R:Author, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig

II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A:Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <MAT>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol

C:Keywords: glycoprotein

F:1-32/Domain: EGF homology (fragment) <EG1>

F:40-79/Domain: discoidin I amino-terminal homology <DN1>

F:82-239/Domain: discoidin I amino-terminal homology <DN2>

F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.3%; Score 635; DB 2; Length 401;

Best Local Similarity 37.9%; Pred. No. 1.1e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLIGHLAGMSTFLVYSNK-----CQPLGMAHGHIRDFQITASQY-- 1149

Db 50 ECQVTDSDHRG--DVFIQYICKPLGVGHGHCETCTSPLMQGTGAIADQSISASSMHLG 107

QY 1150 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGIKTQGARQKFS 1201

Db 108 FMGLQWAPLAPLHQTGIVNAWTSNGYDKNP--WIQVNLARKMWVTGVVTOGASRAGSA 165

QY 1202 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGIGKNIENPPIIARVIRLHPH 1261

Db 166 EYUKTFKVAYSTDGRQFQIQVAGSGDKIFIGNVNSGLKINLFDPLETQVRLVPII 225

QY 1262 YSIRSTRMLMGLCNSCMPGLMESKAISDAQITASSYFTN---MPATWSPSKARLHL 1318

Db 226 CHRGTCLRFELLGCELNGCTEPLGLKNTIPNKQITASSYKWTGLSAPSPFPYARLDN 285

QY 1319 QGRSNARWQVNNPKWQLQVDQKTKMVTGVTQGVKSLTSMYKEFLISSQDGHQWT 1378

Db 286 QGKFNWATQTNASAWLQDLGSKRVTGIIITGARDFGHIGIYVAAYVAGDDGVVWT 345

QY 1379 LFFQNG--KVKVFGNQDSFTPVVNSLDPLRLYLRHPQSVVHQIALMEVLGC 1432

Db 346 EYKDPGASESKIPFGNMNNSHKKNIFFETPFQARFVRIQPVAMHNRIITLVELLGC 401

RESULT 14

S74211

PAS-6/7 protein precursor - bovine

N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 H

C:Species: Bos primigenius taurus (cattle)

C>Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002

C:Accession: S74211; S78114; S24181; S65138; G48394

R:Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob

A:Reference number: S74211; MUID:97008954; PMID:8856064

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <HVA>

A:Cross-references: EMBL:X91895; NID:G1632778; PID:CAA62997.1; PID:G1632779

A:Accession: S78114

A:Molecule type: protein

A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427

R:Kim, D.H.; Kanno, C.; Mizokami, Y.

Biochim. Biophys. Acta 1122, 203-211, 1992

A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from

A:Reference number: S23926; MUID:92353107; PMID:1643094

A:Accession: S24181

A:Molecule type: protein

A:Residues: 383-394 <KIM>

R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal,

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 27-427 <AOK>

R:Author, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-p

II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A:Status: preliminary

A:Molecule type: protein

A:Residues: 233-246 <MAT>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homo

C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-427/Product: PAS-6/7 protein #status experimental <MAT>

F:24-58/Domain: EGF homology <EG1>

F:66-105/Domain: EGF homology <EG2>

F:108-265/Domain: discoidin I amino-terminal homology <DN1>

F:269-427/Domain: discoidin I amino-terminal homology <DN2>

F:234-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted

F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental

F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental

F:59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Query Match 8.3%; Score 635; DB 2; Length 427;

Best Local Similarity 37.9%; Pred. No. 1.2e-34;

Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 1105 ECLIGHLAGMSTFLVYSNK-----CQPLGMAHGHIRDFQITASQY-- 1149

Db 76 ECQVTDSDHRG--DVFIQYICKPLGVGHGHCETCTSPLMQGTGAIADQSISASSMHLG 133

QY 1150 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGIKTQGARQKFS 1201

Db 134 FMGLQWAPLAPLHQTGIVNAWTSNGYDKNP--WIQVNLARKMWVTGVVTOGASRAGSA 191

QY 1202 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGIGKNIENPPIIARVIRLHPH 1261

Db 192 EYUKTFKVAYSTDGRQFQIQVAGSGDKIFIGNVNSGLKINLFDPLETQVRLVPII 251

QY 1262 YSIRSTRMLMGLCNSCMPGLMESKAISDAQITASSYFTN---MPATWSPSKARLHL 1318

Db 252 CHRGTCLRFELLGCELNGCTEPLGLKNTIPNKQITASSYKWTGLSAPSPFPYARLDN 311

QY 1319 QGRSNARWQVNNPKWQLQVDQKTKMVTGVTQGVKSLTSMYKEFLISSQDGHQWT 1378

Db 312 QGKFNWATQTNASAWLQDLGSKRVTGIIITGARDFGHIGIYVAAYVAGDDGVVWT 371

QY 1379 LFFQNG--KVKVFGNQDSFTPVVNSLDPLRLYLRHPQSVVHQIALMEVLGC 1432

Db 372 EYKDPGASESKIPFGNMNNSHKKNIFFETPFQARFVRIQPVAMHNRIITLVELLGC 427

RESULT 15

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25945

R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Mutha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986
A;Title: A large region (approx 95 kDa) of human factor VIII is dispensable for in vitro
A;Reference number: A25945; MUID:86287369; PMID:3016730
A;Accession: A25945
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-869 <TOO>
C;Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase
Query Match 7.1%; Score 543; DB 2; Length 869;
Best Local Similarity 75.2%; Pred. No. 5.4e-28;
Matches 100; Conservative 14; Mismatches 19; Indels 0; Gaps 0;
QY 746 PPVLKRRHOREITRTTLQSDQBEIDYDDTISVEMKKEDFDIYDDENQSPRSFQKTRHYF 805
DB 737 PPVLRRHQDISLPTRPQPEEDKMDYDDIFSTETKGEDEDIYCEDENQDPRSFQKTRHYF 796
QY 806 IAAVRLMDYGMSSPHVLRNRAQSGVFPQKKVVVFQFTDGSFTQPLYRGELNEHLGLL 865
DB 797 IAAVEQLMDYGMSESPRALNRNRAQNGEVRPKKVVFRERADGSGFTNPSYRGELNKLGLL 856
QY 866 GPYIRAEVEDNIM 878
DB 857 GPYIRAEVEDNIM 869

Search completed: December 9, 2003, 17:08:39
Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2003, 16:57:27 ; Search time 18 Seconds

(without alignments)
3756.911 Million cell updates/sec

Title: US-10-047-257-1

Perfect score: 7691

Sequence: 1 ATRRYVLGAVELSDYMQSD.....VVHQIALRMEVLGCAQDLY 1438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	7227	94.0	2351	1	FA8_HUMAN
2	6227	81.0	2133	1	FA8_PIG
3	6192	80.5	2319	1	FA8_MOUSE
4	2389.5	31.1	2224	1	FA5_HUMAN
5	2372	30.8	2211	1	FA5_BOVIN
6	2357.5	30.7	2258	1	FA5_PIG
7	1727.5	22.5	1055	1	CERU_HUMAN
8	1684.5	21.9	1059	1	CERU_RAT
9	1635.5	21.3	1062	1	CERU_MOUSE
10	663	8.6	427	1	MFGM_RAT
11	657	8.5	463	1	MFGM_MOUSE
12	650	8.5	409	1	MFGM_PIG
13	635	8.3	427	1	MFGM_BOVIN
14	588	7.6	387	1	MFGM_HUMAN
15	469.5	6.1	931	1	NRP2_HUMAN
16	464.5	6.0	925	1	NRP2_RAT
17	462.5	6.0	931	1	NRP2_MOUSE
18	458.5	6.0	914	1	NRP1_CHICK
19	451.5	5.9	922	1	NRP1_RAT
20	446.5	5.8	923	1	NRP1_MOUSE
21	443	5.8	928	1	NRP1_XENLA
22	429.5	5.6	923	1	NRP1_HUMAN
23	306.5	4.0	3133	1	HMCT_BOMMO
24	266	3.5	764	1	CPX2_MOUSE
25	261	3.4	756	1	CPX2_HUMAN
26	260.5	3.4	280	1	XLRI_FUGRU
27	247	3.2	224	1	XLRI_MOUSE
28	243	3.2	224	1	XLRI_HUMAN
29	223	2.9	722	1	CPXM_MOUSE
30	217	2.8	734	1	CPXM_HUMAN
31	211	2.7	1331	1	CTA2_HUMAN
32	211	2.7	1284	1	NRX4_DROME
33	208	2.7	1310	1	CTA4_MOUSE

34 206 2.7 1308 1 CTA4_HUMAN
35 198.5 2.6 854 1 DDR2_MOUSE
36 194 2.5 1288 1 CTA3_HUMAN
37 192 2.5 913 1 DDR1_HUMAN
38 185.5 2.4 855 1 DDR2_HUMAN
39 185.5 2.4 910 1 DDR1_RAT
40 185.5 2.4 911 1 DDR1_MOUSE
41 183 2.4 1384 1 CTA1_HUMAN
42 171 2.2 1381 1 CTA1_RAT
43 171 2.2 1385 1 CTA1_MOUSE
44 149.5 1.9 578 1 ASO_TOBAC
45 148 1.9 622 1 YAK8_SCHPO

Q9C0A0 homo sapien
Q62371 mus musculus
Q9B766 homo sapien
Q08345 h epithelia
Q16832 homo sapien
Q63474 rattus norv
Q03146 mus musculus
P78557 homo sapien
P97846 rattus norv
O54991 mus musculus
Q40588 nicotiana t
Q09920 schizosacch

ALIGNMENTS

RESULT 1
ID_FAB_HUMAN STANDARD; PRT; 2351 AA.
AC P00451;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component)
DE (Antithemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=86081164; PubMed=3935400;
RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
RA Hartog K., Kuo C.H., Maslary F.R., Merryweather J.P., Najarian R.,
RA Pachl C., Potter S.J., Puma J., Quiroga M., Rali L.B., Randolph A.,
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
RA Nordfang O., Szaban M.;
RT "Characterization of the polypeptide composition of human factor
RT VIII:C and the nucleotide sequence and expression of the human kidney
RT cDNA."
RL DNA 4:333-349(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RT "Expression of active human factor VIII from recombinant DNA clones."
RL Nature 312:330-337(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061550; PubMed=6438528;
RA Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,
RA Hewick R.M.;
RT "Molecular cloning of a cDNA encoding human antithemophilic factor."
RL Nature 312:342-347(1984).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93265012; PubMed=1303178;
RA Gitschier J., Wood W.I.;
RT "Sequence of the exon-containing regions of the human factor VIII
RT gene."
RL Hum. Mol. Genet. 1:199-200(1992).
RN [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.

RX MEDLINE=91093266; PubMed=1898735;
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeest M.P.,
 RA Mertens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor."; *J. Biol. Chem.* 266:740-746(1991).
 RL J. Biol. Chem. 266:740-746(1991).
 RN [7]
 RP SUTLATION.
 RP MEDLINE=92207952; PubMed=1554716;
 RX Pittman D.D., Wang J.H., Kaufman R.J.;
 RA "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII."; *Biochemistry* 31:3315-3325(1992).
 RL Biochemistry 31:3315-3325(1992).
 RN [8]
 RP STRUCTURE BY NMR OF 2322-2343.
 RP MEDLINE=95200924; PubMed=7893714;
 RX Gilbert G.E., Baleja J.D.;
 RA "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy."; *Biochemistry* 34:3022-3031(1995).
 RL Biochemistry 34:3022-3031(1995).
 RN [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=91221499; PubMed=1902642;
 RX Gitschier J.;
 RA "The molecular basis of hemophilia A."; *Ann. N.Y. Acad. Sci.* 614:89-96(1991).
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=89088506; PubMed=2491949;
 RX White G.C. II, Shoemaker C.B.;
 RA "Factor VIII gene and hemophilia A."; *Blood* 73:11-12(1989).
 RL Blood 73:11-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMA.
 RP MEDLINE=95245332; PubMed=7728145;
 RX Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RA "Molecular etiology of factor VIII deficiency in hemophilia A."; *Hum. Mutat.* 5:1-22(1995).
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT HEMA GLN-2326.
 RP MEDLINE=86235434; PubMed=3012775;
 RX Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
 RA "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic."; *Science* 232:1415-1416(1986).
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT HEMA PRO-2135.
 RP MEDLINE=88096539; PubMed=3122181;
 RX Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
 RA "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences."; *Nucleic Acids Res.* 15:9797-9805(1987).
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT HEMA GLN-2228.
 RP MEDLINE=88191889; PubMed=2833855;
 RX Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides."; *Am. J. Hum. Genet.* 42:718-725(1988).
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT HEMA GLY-291.
 RP MEDLINE=88220354; PubMed=2835904;
 RX Yousoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene."; *Am. J. Hum. Genet.* 42:867-871(1988).
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT HEMA CYS-1708.
 RP MEDLINE=89274393; PubMed=2499363;
 RX O'Brien D.P., Tuddenham E.G.;

RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A."; *Blood* 73:2117-2122(1989).
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT HEMA CYS-391.
 RP MEDLINE=90001543; PubMed=2506948;
 RX Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
 RA "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule."; *Blood* 74:1612-1617(1989).
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT HEMA LEU-189.
 RP MEDLINE=90057680; PubMed=2510835;
 RX Chan V., Chan T.K., Tong T.M., Todd D.;
 RA "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A."; *Blood* 74:2688-2691(1989).
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT HEMA LEU-2326.
 RP MEDLINE=89197216; PubMed=2495245;
 RX Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RA "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene."; *Hum. Genet.* 81:335-338(1989).
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HEMA HIS-391.
 RP MEDLINE=89264602; PubMed=2498882;
 RX Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine)."; *Proc. Natl. Acad. Sci. U.S.A.* 86:4277-4281(1989).
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT HEMA CYS-1708.
 RP MEDLINE=90105723; PubMed=2104766;
 RX Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A."; *Blood* 75:384-389(1990).
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS HEMA GLN-2228 AND LEU-2326.
 RP MEDLINE=90123183; PubMed=2105106;
 RX Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudou F., Carbonara A.;
 RA "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent."; *Blood* 75:662-670(1990).
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT HEMA CYS-391.
 RP MEDLINE=90329422; PubMed=1973901;
 RX Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RA "CRM haemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site."; *Br. J. Haematol.* 75:73-77(1990).
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS HEMA PHE-1699 AND CYS-1708.
 RP MEDLINE=90152691; PubMed=2105906;
 RX Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA."; *Genomics* 6:65-71(1990).
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS HEMA CYS-1728 AND ASP-1941.
 RP MEDLINE=90169988; PubMed=2106480;
 RX Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";

Query Match				94.0%; Score 7227; DB 1; Length 2351;			
Best Local Similarity				61.6%; Pred. No. 0;			
Matches 1437; Conservative				0; Mismatches 1; Indels 894; Gaps 1;			
QY	1	ATRYXILGAVELSDWYQSDLGELPVDARPPRPVKSPFNTSVVYKKTILFVEFTVHLFN	60				
DB	20	ATRYXILGAVELSDWYQSDLGELPVDARPPRPVKSPFNTSVVYKKTILFVEFTVHLFN	79				
QY	61	IAKPRPPMGLGPTIOAEVYDVTIILKNWASHPUSLHAGVSVYKASGAEVDDOTSQ	120				
DB	80	IAKPRPPMGLGPTIOAEVYDVTIILKNWASHPUSLHAGVSVYKASGAEVDDOTSQ	139				
QY	121	REKEDDVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLGALLVCR	180				
DB	140	REKEDDVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLGALLVCR	199				
QY	181	EGSLAKETQTLHKPILLFVDFEGKSNHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	240				
DB	200	EGSLAKETQTLHKPILLFVDFEGKSNHSETKNSLMQDRDAASARAWPKOHTVNGYVNR	259				
QY	241	SLPGLICHRKSVVHVIGMGTPEVHSIFLEGHTPLVRNHRQASLEISDITFLTAQTLL	300				
DB	260	SLPGLICHRKSVVHVIGMGTPEVHSIFLEGHTPLVRNHRQASLEISDITFLTAQTLL	319				
QY	301	MDLGQFLFCHISQHDGMEAYVKVDSCEPEEQRLMKNEEAEDYDDDLTDSMDVVRP	360				
DB	320	MDLGQFLFCHISQHDGMEAYVKVDSCEPEEQRLMKNEEAEDYDDDLTDSMDVVRP	379				
QY	361	DDNSPSFIQIRSAVKHPKTHWHYIAAEEEDWDYAPLVAPDDRYSKYSLNNGPQIRG	420				
DB	380	DDNSPSFIQIRSAVKHPKTHWHYIAAEEEDWDYAPLVAPDDRYSKYSLNNGPQIRG	439				
QY	421	RKYKVRFMAYTDETFKTRAIQHEGILGPLLYGVEGDTLLIIKNOQASRPNIYPHGI	480				
DB	440	RKYKVRFMAYTDETFKTRAIQHEGILGPLLYGVEGDTLLIIKNOQASRPNIYPHGI	499				
QY	481	TDVRPLYSRLPKGVKHLKDFPILPGEIFIKYKVTVTVEDGPTKSDPRCLTRYSSVFNME	540				
DB	500	TDVRPLYSRLPKGVKHLKDFPILPGEIFIKYKVTVTVEDGPTKSDPRCLTRYSSVFNME	559				
QY	541	RLASGLIGLILLICYKESVDQRNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG	600				
DB	560	RLASGLIGLILLICYKESVDQRNQIMSDKRNVLFSVFDENRSWYLTENIQRLPNPAG	619				
QY	601	VOLEDPEFOASNIWHSNGVVFDSLOLSVCLHBAVWYIILSIGAOTDELSPESGTEKH	660				
DB	620	VOLEDPEFOASNIWHSNGVVFDSLOLSVCLHBAVWYIILSIGAOTDELSPESGTEKH	679				
QY	661	KWYEDTLTLFPFSGTVMFMSNPGWLILGCHNSDFRNGMTALLKVSCKNTGDYVE	720				
DB	680	KWYEDTLTLFPFSGTVMFMSNPGWLILGCHNSDFRNGMTALLKVSCKNTGDYVE	739				
QY	721	DSYEDISAYILLSKNAIEPRSF	742				
DB	740	DSYEDISAYILLSKNAIEPRSF	799				
QY	743	-----	742				
DB	800	KIQNVSSDLLMLLRQSPHPHGLSLDLQAEKAYETSDPSPCAIDSNNSLSEMTWFRPQ	859				
QY	743	-----	742				
DB	860	LHHSQDMVFTPESGLQRLNEKLGTTAATLKLDFKVSSTSNLJSTIPSDNLAAGTDN	919				
QY	743	-----	742				
DB	920	TSSLGPPSPVHYDSQDITTLFGKKSSPLTBSGGPLSLEENNDKLLSGLMSQESSW	979				
QY	743	-----	742				
DB	980	GKNVSTESGRFLFKGKAHPALLTKDNALFKVISLTKNTKSNNSATNRKTHIDGSL	1039				
QY	743	-----	742				
DB	1040	LIENSFVWQNILESDETEFKKVTPLIHDRMLDKNATALRLNHSNKTTSKMEMVQOK	1099				
QY	743	-----	742				
DB	1100	KEGPIDDAQNPDMSPFKMLFLPESARWQRTQTHGKNSLNSGQSPSPKQLVSLGPEKSV	1159				
QY	743	-----	742				
DB	1160	QNFELSEKNKVVVGKGBFTKDVGLKEMVFPSSRNFLTLNLDNLHNHNTNQEKKIQBEIEK	1219				
QY	743	-----	742				
DB	1220	KETLIOENVVLPOIHTVTGTONPKMKNLFLLLSTRQNVGSDGAYAPVLQDFRSLNDSTNR	1279				
QY	743	-----	742				
DB	1280	TKKHTAHFSKKGEEENLEGNGQTKQIWEKYACTTRISPTNSQNFVTORESKEALKQFRL	1339				
QY	743	-----	742				
DB	1340	PLEETELEKRIIVDDTSTQSKNMKHLTPSLTQIDYNEKEKGAITQSPSLDCLTRSHSI	1399				
QY	743	-----	742				
DB	1400	PQANRSLPIAKVSSPSPRIPIYLRVLFQDNSSHLPAASRYKKSQGVQESSHFLQSAKK	1459				
QY	743	-----	742				
DB	1460	NNLSLAILLEMTQDQREVSGSLGTSATNSVYKVENTVLPKPDLLPKTSCKVELLPKVHI	1519				
QY	743	-----	742				
DB	1520	YOKDLFPPTETSGSPGHLDLVEGSLIQGTGAIKWNEANRPGKVPFLRVATESSAKTPSK	1579				
QY	743	-----	742				
DB	1580	LLDPLAWDNHYQTQPKBEWKSQESKPEKTAFFKDDTILSLNACSNHAAINEGONKP	1639				
QY	743	-----	742				
DB	1640	ELEVWAKQGRTERLCSQNPFPVLKXHQREITATLQSDQEEIDYDDTISVEMKKEFDIY	1699				
QY	787	DEDEQSPRSFOKTRHPIAVERLWYGMSSSPHVLNRAQSGSVPOFKKVVQFQFTD	846				
DB	1700	DEDEQSPRSFOKTRHPIAVERLWYGMSSSPHVLNRAQSGSVPOFKKVVQFQFTD	1759				
QY	847	GSFTQPLVRGBELNEHLGLLGPYIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEORQGA	906				
DB	1760	GSFTQPLVRGBELNEHLGLLGPYIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEORQGA	1819				
QY	907	EPRKNFVKPNETKTYFWKQVHHMPTKDEFCCKAWAYFSDVLEKDVHSGSLIGPLLVCHT	966				
DB	1820	EPRKNFVKPNETKTYFWKQVHHMPTKDEFCCKAWAYFSDVLEKDVHSGSLIGPLLVCHT	1879				
QY	967	NFLNPAHQROVTVQSFALFFTFIDETKSWYTFENMERNCRAPCNTOEMDDPTFKENYRHA	1026				
DB	1880	NFLNPAHQROVTVQSFALFFTFIDETKSWYTFENMERNCRAPCNTOEMDDPTFKENYRHA	1939				
QY	1027	INGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEEYKMAIYNLYP	1086				
DB	1940	INGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIFSGHVFTVRKKEEYKMAIYNLYP	1999				
QY	1087	GVFETVEMLPKAGIRVVECLIGELHAGMSTFLVYSNKQCTPLGMAASHIRDFQITAS	1146				
DB	2000	GVFETVEMLPKAGIRVVECLIGELHAGMSTFLVYSNKQCTPLGMAASHIRDFQITAS	2059				
QY	1147	GOYGQWAPKLARLHYSGINSINAWSTKPSWIKVDLAPMIITHGKTQARQKFSLSYISQ	1206				
DB	2060	GOYGQWAPKLARLHYSGINSINAWSTKPSWIKVDLAPMIITHGKTQARQKFSLSYISQ	2119				
QY	1207	FIIMYSLDGKKQWTVRGNSTGTLMVFFGNVDSSGKKNIFNPPIIARVIRLHPHYHSIRS	1266				

Db 2120 FILMSLDGKWTYRGNSTGLMVFEGNVDSGKHNIPNPIIARYITLPHPTYSIRS 2179
 Qy 1267 TLRMELMGCDLNSCSPLGMSKASDAQITASSYFTNMFATWSPSKARHLQGRSNMR 1326
 Db 2180 TLRMELMGCDLNSCSPLGMSKASDAQITASSYFTNMFATWSPSKARHLQGRSNMR 2239
 Qy 1327 PQVNNPKWLVQDFQKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 1386
 Db 2240 PQVNNPKWLVQDFQKTMKVTVGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
 Qy 1387 KVFQGNQDSPTVNSLDPLRLRYLRHQPQSWHQIALRMEVLGCEAQDLY 1438
 Db 2300 KVFQGNQDSPTVNSLDPLRLRYLRHQPQSWHQIALRMEVLGCEAQDLY 2351

RESULT 2

FA8_PIG STANDARD; PRT; 2133 AA.
 AC P12263; Q95243;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Coagulation factor VIII precursor (Procoagulant component).
 GN F8 OR C8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A. A.
 RA Healey J.F., Lubin I.M., Lollar P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 705-1573 FROM N.A.
 RX MEDLINE=66287369; PubMed=3016730;
 RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
 KAUFMAN R.J.;
 RT "A large region (approximately equal to 95 kDa) of human factor VIII
 is dispensable for in vitro procoagulant activity."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942 (1986).
 RN [3]
 RP SEQUENCE OF 392-759 FROM N.A.
 RX MEDLINE=9417260; PubMed=7510693;
 RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
 RT "Elimination of a major inhibitor epitope in factor VIII."
 RL J. Biol. Chem. 269:8639-8641 (1994).
 CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
 CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
 CC ACTIVATED FORM, FACTOR XA.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC -----
 DR EMBL; U49517; AB06705.1; -;
 DR PIR; A25945; A25945.
 DR PIR; T42763; T42763.
 DR HSP; P00451; ICFG.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00394; Cu-oxidase_3.
 DR Pfam; PF00754; F5_P8Type_C_2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS01285; FA58C_1; 2.

DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00222; FA58C_3; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
 KW Signal; Glycoprotein; Sulfation.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2133 COAGULATION FACTOR VIII.
 FT DOMAIN 20 357 F5/8 TYPE A 1.
 FT DOMAIN 20 139 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 399 730 F5/8 TYPE A 2.
 FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 760 1599 B.
 FT DOMAIN 1495 1822 F5/8 TYPE A 3.
 FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1669 1822 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 1822 1970 F5/8 TYPE C 1.
 FT DOMAIN 1975 2127 F5/8 TYPE C 2.
 FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
 FT SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
 FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
 FT DISULFID 173 199 PROBABLE.
 FT DISULFID 547 573 PROBABLE.
 FT DISULFID 1633 1659 PROBABLE.
 FT DISULFID 1822 1970 BY SIMILARITY.
 FT DISULFID 1975 2127 BY SIMILARITY.
 FT CARBOHYD 233 233 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 929 929 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 985 985 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1025 1025 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1111 1111 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1181 1181 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1208 1208 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1245 1245 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1265 1265 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1335 1335 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1408 1408 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1611 1611 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 1919 1919 N-LINKED (GLCNAC). (POTENTIAL).
 FT CONFLICT 713 713 N -> M (IN REF. 2).
 FT CONFLICT 734 734 I -> T (IN REF. 2).
 FT CONFLICT 792 792 G -> Q (IN REF. 2).
 FT CONFLICT 1133 1133 E -> F (IN REF. 2).
 FT CONFLICT 1191 1191 I -> L (IN REF. 2).
 FT CONFLICT 1209 1209 R -> F (IN REF. 2).
 FT CONFLICT 1437 1437 C -> G (IN REF. 2).
 FT CONFLICT 1456 1456 F -> R (IN REF. 2).
 FT CONFLICT 1539 1539 F -> R (IN REF. 2).
 FT CONFLICT 1546 1546 Q -> N (IN REF. 2).
 SQ SEQUENCE 2133 AA; 239304 MW; 152BBA6997F570DA CRC64;

Query Match

81.0%; Score 6227; DB 1; Length 2133;

Best Local Similarity 57.3%; Pred. No. 0;

Matches 1212; Conservative 106; Mismatches 119; Indels 678; Gaps 3;

Qy 1 ATRRYLGAVELSWDYMQSD-LGELFVDARPPRPVKSPFPNTSVYKKTILFVEVTVHLF 59
 Db 20 AIRRYLGAVELSWDYRQSELRRLSHVDTRFPATAPGALPLGPSVLVYKKTIVFEVTDQLF 79
 Qy 60 NIAXRPPPMWGLGPTIAEYVDVTVTLKNWASHPVSHANGSVYKASGAEYDDOTS 119
 Db 80 SVARPPPMWGLGPTIAEYVDVTVTLKNWASHPVSHANGSVYKASGAEYDDOTS 139
 Qy 120 QREKEDDKVFGSGSHYVQVQLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVC 179
 Db 140 QREKEDDKVFGSKQTYVQVQLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVC 199

FT	CARBOHYD	259	259	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	601	601	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	880	880	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	958	958	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1015	1015	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1022	1022	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1026	1026	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1076	1076	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1087	1087	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1136	1136	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1161	1161	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1192	1192	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1255	1255	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1268	1268	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1273	1273	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1274	1274	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1302	1302	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1316	1316	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1340	1340	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	1797	1797	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
FT	CARBOHYD	2105	2105	N-LINKED	(GLCNAC.	.	.)	(POTENTIAL).
SQ	SEQUENCE	2319	AA; 266148 MW; FD054DE051DB2A01 CRC664;					
Query Match 80.5%; Score 6192; DB 1; Length 2319;								
Beat Local Similarity 53.4%; Pred. No. 0;								
Matches 1231; Conservative 97; Mismatches 103; Indels 876; Gaps 9								
Qy	1	ATRRYVLGAVELSDWYMQSD-LGELPVDARFPFRVPKSPFNTSVVYKTLFVFTVHLF	59					
Db	20	AIRRYVLGAVELSWNYIQSDLLSVLHSDSRFLPRMSTSPFNTSIMYKTVFVBYKQOLF	79					
Qy	60	NIAPRPPWMLGLPTIQAEVDTVVITLKNASHPVLHVGVSYNKASGABYDDQTS	119					
Db	80	NIAPRPPWMLGLPTIMTEVHDVVITLKNASHPVLHVGVSYNKASGDEYEDQTS	133					
Qy	120	QREKEDKVPFGSHYVQVILKENGPMASDPLCLTYSLSHVDLVKDLNLSGLLGALLVC	176					
Db	140	QREKEDKVPFGESHYVWQVILKENGPMASDPPCLTYSYMSHVDLVKDLNLSGLLGALLVC	198					
Qy	180	REGSLAKETQTLHKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTWGVYN	239					
Db	200	KEGSLKERTOMLYQFVLLFAVDFEGKSWHSETNDSYQOSMDSASADMPQKMTWGVYN	258					
Qy	240	RSIPGLIGCHRSVYHVIIGMGTTEVHSIIFLEHTFLVRHHRQASLEISPTFLTAQTL	298					
Db	260	RSPLGLIGCHRSVYHVIIGMGTTPETIHSIIFLEHTFPVRHHRQASLEISPTFLTAQTL	319					
Qy	300	LMDLQGLFLFCHISSHOHDGMEAYKVVDSCPEEPOLRMK-NNEBAEDYDDDLTDESDMVV	356					
Db	320	LIDLQGLFLFCHISSHKHDGMEAYKVVDSCPEESQWQKNNEMEDYDDDLV-SEMDMF	378					
Qy	359	RFDDNSPSFQIRSVAKKPKTWVHYHTAAEEEDWDYAPLVAPLAPDRSYKSQYLNNQFOR	418					
Db	379	TLDYDSSP-FIQIRSVAKKYPKTIHYIISABEEDWDYAPSVPTSDNGSYKSQYLSNGPGR	437					
Qy	419	IGBKYKKVRFMAYTDETAKTREAIOHESGILGPLLYKEVGDTLLIIIPKQASAPRYNIYPH	478					
Db	438	IGBKYKKVRFIAYTDETAKTRETIOHESGLGLGPLLYKEVGDTLLIIIPKQASAPRYNIYPH	497					
Qy	479	GITDVRPLYSRRLPKGVKHLKDFPLPOEIEFKYKMTVTVBDGPTKSPRCLTRYSSEFVN	538					
Db	498	GITDVSPLHARLPRGIKHVKDLPHPGEIIFKYKMTVTVBDGPTKSPRCLTRYSSEFIN	557					
Qy	539	MERDLASGLIGPLLIICYKESVDQRGNQIMSKRNVILFVSFDENRSYLTENIQRELPN	598					
Db	558	PERDLASGLIGPLLIICYKESVDQRGNQIMSKRNVILFVSFDENQSYLTENIQRELPNA	617					
Qy	599	AGVQLEDEPFOASNTMHSINGVTFDSLOLSVCLHEVAYWILSGAOTDPLSPFSGYTF	658					
Db	618	AKTQPDQFOASNTMHSINGVTFDSLELTVCLHEVAYWILSVGAOTDPLSPFSGYTF	677					

```
QY 659 KKKVYEDTLTLPFSGETVFMENPGILWILGCHNSDPRNRGWTALLKVVSCDKRTGY 718
DB 678 KKKVYEDTLTLPFSGETVFMENPGILWILGCHNSDPRNRGWTALLKVVSCDKRTSDY 737
QY 719 YEDSYEDIKAYLLSKNNALPRSFQ-- 745
DB 738 YBEIYEDITQLVNNVNDPRSFQNTWHPNTRKKKFKVDSIPKNDMEKIBPQFEIAE 797
QY 746 ----- 745
DB 798 MLKVQSVSDMLMLGQSHPTPHGLFLSDGOBAIYEAIHDSHPNAIDSEGPSKVQL 857
QY 746 ----- 745
DB 858 RPESHSEKIVFPQGLQLRNKSLETTIEVKWKGLQVSSLPSNLMTTILSDNLKA 917
QY 746 ----- 745
DB 918 TPEKTDSSGFPDMPVHSSKSLTTFAGKKAISLVGSHVPLNASENSDSNILDSTLMYSQ 977
QY 746 ----- 745
DB 978 ESLPRDNILSIENDRLLEKRRFHGIALTKDNTLFDKNVSLMKTNTKYNHSTTNEKLHTE 1037
QY 746 ----- 745
DB 1038 SPTSISNTDQLDAIKVNSEIQEVTAIHGDTLLGKNSYILRLHMLNLRITSTKQKI 1097
QY 746 ----- 745
DB 1098 FHRKDBDPIQDEBNTIMPFSKWLFLSESNWPKTNGNSLSEHQESPKQLVILMPKK 1157
QY 746 ----- 745
DB 1158 YVKNQSLSEKKNVTEQDFTKNIIGLKDMPHNNMSIFLTLNHNHENGHNQEKNIQE 1217
QY 746 ----- 745
DB 1218 EIEKEALIEKVLPQVHEATGSKNFKDLILGTRQNISLVEHVVPVLQNTISINNSTN 1277
QY 746 ----- 745
DB 1278 TVQIHMEHFEXRKOKETNSEGLVNTKREWVKNYPSQKNITQRSKRALGQFELSTQWLK 1337
QY 746 ----- 745
DB 1338 TINCSTQCIKQIDHSKEMKKFITKSLSDSVIKSTTQTNSSDSHIVKTSAPPPIDLKR 1397
QY 746 ----- 745
DB 1398 SPFQNFHSHVQASSIYDFKTKSSRIQESNNFLKETHKINPNLSLILPNNMFIQDGKFTSP 1457
QY 746 ----- 745
DB 1458 GKSNTNSVYKKRENIIFLKPTLPBESGKIELLPQVSIQEEBILPTETSHGSPGHLNLMK 1517
QY 746 ----- 745
DB 1518 EVFLQIQGPTKWKAKRHGESIKGTBSSKNTRSKLNNHAWDYHYAAQIPKDMWKSKE 1577
QY 746 ----- PPVLKR 751
DB 1578 KSPFIIISIQEDTILSLRPHGNSHSIGANEKQWNPQRETTWYKQGTQRTCSQIPVLKR 1637
QY 752 HOREITRPTLQSDQEBIDDDTISVEMKEDFDIYDENQSPRFSQKTRHYFAAVER 811
DB 1638 HOREL--SAFQSEQATDYDAITIE--TIEDEFDIYSEDIKQGPFSFQKTRHYFAAVER 1694
QY 812 LWDYGMSSSPHVLNRNAQSGSVQPKVVFQFTDGSFTQPLVYRGELNEHLGLGYIRA 871
DB 1695 LWDYGMSTG--HVLNRNYQSDNVQPKVVFQFTDGSFTQPLVYRGELNEHLGLGYIRA 1753
```

```
QY 872 EVEDNIMTFNRQASRPYSFYSSLSYSEDQOQAEPRKPNFKPNETKTYFKWQVHMAP 931
DB 1754 EVEDNIMTFNRQASRPYSFYSSLSYSEKEDQO--GEBPRRNFKPNETKTYFKWQVHMAP 1812
QY 932 TKDEDFCKAWAYFSDVLEKDVHSLIGPLVCHTNTLNPAGRGROVTVQEFALFFTFDE 991
DB 1813 TEDEDFCKAWAYFSDVLEKDVHSLIGPLVCHTNTLNPAGRGROVTVQEFALFFTFDE 1872
QY 992 TKSWYFTENMERNCRAPCNIQMEDPTFKENTRPHAINGYIMDTLPGLVMAQDQIRWYLL 1051
DB 1873 TKSWYFTENMERNCRAPCNIQMEDPTFKENTRPHAINGYIMDTLPGLVMAQDQIRWYLL 1932
QY 1052 SHGSHENTHSIHFGSHVFTVRKKEYSKALNYLYGVFETVEMLSKAGINVECLIGEH 1111
DB 1933 SHGSHENTHSIHFGSHVFTVRKKEYSKALNYLYGVFETVEMLSKAGINVECLIGEH 1992
QY 1112 LHAGMSTLFLVYSNCKQTPLGWASGHIRDFOITASGOYQWAPKRLARLHYSGSINAWSTK 1171
DB 1993 LHAGMSTLFLVYSNCKQTPLGWASGHIRDFOITASGOYQWAPKRLARLHYSGSINAWSTK 2052
QY 1172 EPPSWIKVDLLAPMIHGIKTQARQKFSYISQIIFIMYSLDGKKWQTYRGNSTGTLMV 1231
DB 2053 EPPSWIKVDLLAPMIHGIKTQARQKFSYISQIIFIMYSLDGKKWQTYRGNSTGTLMV 2112
QY 1232 FEGNVDSGKIKHNIENPPIIARYIPLHPTHSIRSTLPMELMGCGLNSCSPMLGSHSKAI 1291
DB 2113 FEGNVDSGKIKHNIENPPIIARYIPLHPTHSIRSTLPMELMGCGLNSCSPMLGSHSKAI 2172
QY 1292 SDAQITASSYFTNMFATWSPSKARLHLOQRSNAMRPQVNNPKEMIQVDFOKTMKVGVTT 1351
DB 2173 SDAQITASSYFTNMFATWSPSKARLHLOQRSNAMRPQVNNPKEMIQVDFOKTMKVGVTT 2232
QY 1352 QGVKSLTSMYKVEFLISSQDGHQWTLFPQNGKVKVQGNQDSTPVPVNSLDPPLLTREY 1411
DB 2233 QGVKSLTSMYKVEFLISSQDGHQWTLFPQNGKVKVQGNQDSTPVPVNSLDPPLLTREY 2292
QY 1412 LRIHPQSMVHOIALRMEVLGCEAODLY 1438
DB 2293 LRIHPQSMVHOIALRMEVLGCEAODLY 2319

RESULT 4
FA5_HUMAN
ID FA5_HUMAN STANDARD; PRT; 2224 AA.
AC P1259; Q14285;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A., AND VARIANT MET-1764.
RX MEDLINE=92232668; PubMed=1567832;
RA Crispe L.D., Moore K.D., Kane W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL Biochemistry 31:3777-3785(1992).
RN [2]_
RP SEQUENCE FROM N.A., AND VARIANT GLU-925.
RX MEDLINE=87260886; PubMed=3110773;
RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
RA Hewick R.M., Kaufman R.J., Mann K.G.;
RT "Complete cDNA and derived amino acid sequence of human factor V.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
RN [3]_
RP SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GLU-925 AND ILE-1285.
RX MEDLINE=88107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
```


RT of internal repeats.";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313665; PubMed=3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RX TISSUE=Fibroblast;
 RA Shen N.I.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes.";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE=94264012; PubMed=8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RP SULFATION.
 RX MEDLINE=90366699; PubMed=2168225;
 RA Hottin G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V.";
 RL Blood 76:946-952(1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE=20052169; PubMed=10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V.";
 RL Nature 402:434-439(1999).
 RN [9]
 RP VARIANT MET-1764.
 RX MEDLINE=95179146; PubMed=7874144;
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;
 RT "A polymorphism in the human coagulation factor V gene.";
 RL Hum. Mol. Genet. 3:2085-2085(1994).
 RN [10]
 RP VARIANT APCR GLN-534.
 RX MEDLINE=94217810; PubMed=8164741;
 RA Bertina R.M., Koelmann B.P.C., Koester T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C.";
 RL Nature 369:64-67(1994).
 RN [11]
 RP VARIANTS ILE-1285 AND ARG-1327.
 RX MEDLINE=96351768; PubMed=8713778;
 RA Lunghi B., Iacovello L., Gemmati D., Dilasio M.G., Castoldi E.,
 RA Pinotti M., Castanan G., Redaelli R., Mariani G., Marchetti G.,
 RA Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 RT association with factor V levels in plasma.";
 RL Thromb. Haemost. 75:45-48(1996).
 RN [12]
 RP VARIANT APCR GLN-334, AND VARIANT LYS-513.
 RX MEDLINE=98122763; PubMed=9454741;
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
 RL Blood 91:1135-1139(1998).
 RN [13]
 RP VARIANT APCR THR-334.
 RX MEDLINE=98122764; PubMed=9454742;
 RA William D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C.";
 RL Blood 91:1140-1144(1998).
 RN [14]
 RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RP ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RP ILB-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 RT of human genes.";
 RL Nat. Genet. 22:231-238(1999).
 RN [15]
 RP ERRATUM.
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373(1999).
 CC -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia
 CC [MIM:227400], an hemorrhagic diastasis.
 CC -!- DISEASE: Defects in F5 are the cause of resistance to activated
 CC protein C (APCR) [MIM:189055], a form of thrombophilia. The APCR
 CC mutation is found in about 5% of the population which suggest that
 CC a slight thrombotic tendency may confer some advantage in fetal
 CC implantation.
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L32779; AAB59401.1; JOINED.
 DR EMBL; L32755; AAB59401.1; JOINED.
 DR EMBL; L32756; AAB59401.1; JOINED.
 DR EMBL; L32757; AAB59401.1; JOINED.
 DR EMBL; L32758; AAB59401.1; JOINED.
 DR EMBL; L32759; AAB59401.1; JOINED.
 DR EMBL; L32760; AAB59401.1; JOINED.
 DR EMBL; L32761; AAB59401.1; JOINED.
 DR EMBL; L32762; AAB59401.1; JOINED.
 DR EMBL; L32763; AAB59401.1; JOINED.
 DR EMBL; L32764; AAB59401.1; JOINED.
 DR EMBL; L32765; AAB59401.1; JOINED.
 DR EMBL; L32766; AAB59401.1; JOINED.
 DR EMBL; L32767; AAB59401.1; JOINED.
 DR EMBL; L32768; AAB59401.1; JOINED.
 DR EMBL; L32769; AAB59401.1; JOINED.
 DR EMBL; L32770; AAB59401.1; JOINED.
 DR EMBL; L32771; AAB59401.1; JOINED.
 DR EMBL; L32772; AAB59401.1; JOINED.

Db	1699	AVQNSSTVYVHATERSGPGSACRAWAYANVPEKDIHSLGLICQKILHK	1758
Qy	972	AHGQVTVQEPALFTTFIDETKSYFTENMERNCRAPNIQMEDPTFKENYRFAHNGYI	1031
Db	1759	DSNPVDNREVFLLPMTDEKKSWEYKRSRWR-----LTSSEMKSHPEFALNGMI	1812
Qy	1032	MOTLPGLVMAQDRIWRLLSGNSNETHSHFSGHVFVTKKERYKVALNLYPGVPT	1091
Db	1813	Y-SLFGKLYQDQWVRLHLLNITGSGQDIHVHVFHGTLLNENKQKHQGLVMPPLPGSFT	1871
Qy	1092	VEMLPSKAGIWRVECLIGEHFHAGNSTLFLVYSNKCOTPLGMAHGHRDFOITASGOYQ	1151
Db	1872	LEMKASKGWLINTEVGENQAGNQTPFLINDRDCRMPGMLSTGIISDSQIKASERLGY	1931
Qy	1152	WAPKLARLHYSGINAWSTKE---PFS---WIKVLLAPMIHIGIKTOGARQKSSLIYS	1205
Db	1932	MEPLRLANNGSGYNASWEKLABFASKPTQVDMQKEVITGTQOGAKHYLKSCTYT	1991
Qy	1206	QFLIMYLDGKKWOTYRGNSTCTLVVFCGNVDSSGKHNFPPIIARIYRHLPHYSIR	1265
Db	1992	EYVAYSSNQINQWIKFGKSTNRVMFNGNSDRSTIKENQFPDPIVARIYRISPRAYNR	2051
Qy	1266	STLRMLMGCDLNSGSMPLGMSKAYSQAQITASSYFTNMFAT-WSPSKARHLQGRNA	1324
Db	2052	PLRLLEQCEVNGCSTPLGMEKGIENKQITASSFKKSGWGDYWEPPFARLNAQGRVNA	2111
Qy	1325	WRPQVNPKEWLOVDFQKMTGVTGTQGVKLLTSWYKEFLISSQDGHQWTLFFQNG	1384
Db	2112	WQAKANNKQWLEIDLLKTKITAITGTQCKSLSEMYKSTYHSEGGVWPKRYLKS	2171
Qy	1385	KV--KVFQGNDSFTPVNSLDPPLTVRLYRHPQSVHQAIRMEVLGCEAQDLY	1438
Db	2172	SWDKLFGNTKTKGHVKRQFPPIISIRFIRVPTKWNQSIIRLELFC---DIY	2224
RESULT 5			
FA5_BOVIN	STANDARD;	PRT;	2211 AA.
AC	Q28J07;	Q28108;	
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	15-SEP-2003	(Rel. 42, Last annotation update)	
DE	Coagulation factor V precursor (Activated protein C cofactor).		
GN	F5.		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=92147638; PubMed=1737753;		
RA	Guinco E.R., Esmen C.T., Mann K.G., Macgillivray R.T.;		
RL	"The complete cDNA sequence of bovine coagulation factor V.";		
RL	J. Biol. Chem. 267:2971-2978(1992).		
CC	-1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES		
CC	WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.		
CC	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light		
CC	chain, noncovalently bound. The interaction between the two chains		
CC	is calcium-dependent.		
CC	-1- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14		
CC	AA repeats.		
CC	-1- PM: Thrombin activates factor V proteolytically to the active		
CC	cofactor, factor Va (formation of a heavy chain at the N-		
CC	terminus and a light chain at the C-terminus).		
CC	-1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND		
CC	ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).		
CC	-1- SIMILARITY: Contains 3 F5/8 type A domains.		
CC	-1- SIMILARITY: Contains 2 F5/8 type C domains.		
CC	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M81440; AAA30512.1; --
EMBL; M81441; AAA30513.1; --
PIR; A42580; KFBOS.
DR HSSP; P12259; 1CZT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58 C.
PFam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5 F8 type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C.1; 2.
DR PROSITE; PS01286; FA58C.2; 2.
DR PROSITE; PS00022; FA58C.3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL 1 28
FT CHAIN 29 2211
FT CHAIN 29 741
FT PEPTIDE 742 1564
FT
FT CHAIN 1565 2211
FT DOMAIN 30 327
FT DOMAIN 30 193
FT DOMAIN 203 327
FT DOMAIN 348 686
FT DOMAIN 348 525
FT DOMAIN 535 686
FT DOMAIN 696 1564
FT DOMAIN 1124 1151
FT REPEAT 1124 1137
FT REPEAT 1138 1151
FT REPEAT 1188 1453
FT
FT REPEAT 1188 1196
FT REPEAT 1197 1205
FT REPEAT 1206 1214
FT REPEAT 1215 1223
FT REPEAT 1224 1232
FT REPEAT 1233 1241
FT REPEAT 1242 1250
FT REPEAT 1251 1259
FT REPEAT 1260 1268
FT REPEAT 1269 1277
FT REPEAT 1278 1286
FT REPEAT 1287 1295
FT REPEAT 1296 1304
FT REPEAT 1305 1313
FT REPEAT 1314 1322
FT REPEAT 1323 1331
FT REPEAT 1332 1340
FT REPEAT 1341 1349
FT REPEAT 1350 1358
FT REPEAT 1359 1367
FT REPEAT 1368 1376
FT REPEAT 1377 1385
FT REPEAT 1386 1394
FT REPEAT 1395 1403
FT REPEAT 1404 1412
FT REPEAT 1413 1421
FT REPEAT 1422 1430
FT REPEAT 1431 1439
FT REPEAT 1440 1444
FT REPEAT 1445 1453
FT DOMAIN 1569 1738
FT DOMAIN 1738 1738

POTENTIAL.
COAGULATION FACTOR V.
HEAVY CHAIN (BY SIMILARITY).
ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
LIGHT CHAIN (BY SIMILARITY).
F5/8 TYPE A 1.
PLASTOCYANIN-LIKE 1.
PLASTOCYANIN-LIKE 2.
F5/8 TYPE A 2.
PLASTOCYANIN-LIKE 3.
PLASTOCYANIN-LIKE 4.
B.
2 X 14 AA TANDDEM REPEATS.
1-1.
1-2.
30 X 9 AA APPROXIMATE TANDDEM REPEATS OF
[AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
2-1.
2-2.
2-3.
2-4.
2-5.
2-6.
2-7.
2-8.
2-9.
2-10.
2-11.
2-12.
2-13.
2-14.
2-15.
2-16.
2-17.
2-18.
2-19.
2-20.
2-21.
2-22.
2-23.
2-24.
2-25.
2-26.
2-27.
2-28.
2-29 (PARTIAL).
2-30.
F5/8 TYPE A 3.
PLASTOCYANIN-LIKE 5.


```
1285 1293 1294 1302 1311 1320 1329 1338 1347 1356 1365 1374 1383 1392 1401 1410 1419 1428 1437 1446 1455 1464 1473 1482 1491 1500 1509 1518 1527 1536 1545 1554 1563 1572 1581 1590 1599 1608 1617 1626 1635 1644 1653 1662 1671 1680 1689 1698 1707 1716 1725 1734 1743 1752 1761 1770 1779 1788 1797 1806 1815 1824 1833 1842 1851 1860 1869 1878 1887 1896 1905 1914 1923 1932 1941 1950 1959 1968 1977 1986 1995 2004 2013 2022 2031 2040 2049 2058 2067 2076 2085 2094 2103 2112 2121 2130 2139 2148 2157 2166 2175 2184 2193 2202 2211 2220 2229 2238 2247 2256 2265 2274 2283 2292 2301 2310 2319 2328 2337 2346 2355 2364 2373 2382 2391 2400 2409 2418 2427 2436 2445 2454 2463 2472 2481 2490 2500 2509 2518 2527 2536 2545 2554 2563 2572 2581 2590 2600 2609 2618 2627 2636 2645 2654 2663 2672 2681 2690 2700 2709 2718 2727 2736 2745 2754 2763 2772 2781 2790 2800 2809 2818 2827 2836 2845 2854 2863 2872 2881 2890 2900 2909 2918 2927 2936 2945 2954 2963 2972 2981 2990 3000 3009 3018 3027 3036 3045 3054 3063 3072 3081 3090 3100 3109 3118 3127 3136 3145 3154 3163 3172 3181 3190 3200 3209 3218 3227 3236 3245 3254 3263 3272 3281 3290 3300 3309 3318 3327 3336 3345 3354 3363 3372 3381 3390 3400 3409 3418 3427 3436 3445 3454 3463 3472 3481 3490 3500 3509 3518 3527 3536 3545 3554 3563 3572 3581 3590 3600 3609 3618 3627 3636 3645 3654 3663 3672 3681 3690 3700 3709 3718 3727 3736 3745 3754 3763 3772 3781 3790 3800 3809 3818 3827 3836 3845 3854 3863 3872 3881 3890 3900 3909 3918 3927 3936 3945 3954 3963 3972 3981 3990 4000 4009 4018 4027 4036 4045 4054 4063 4072 4081 4090 4100 4109 4118 4127 4136 4145 4154 4163 4172 4181 4190 4200 4209 4218 4227 4236 4245 4254 4263 4272 4281 4290 4300 4309 4318 4327 4336 4345 4354 4363 4372 4381 4390 4400 4409 4418 4427 4436 4445 4454 4463 4472 4481 4490 4500 4509 4518 4527 4536 4545 4554 4563 4572 4581 4590 4600 4609 4618 4627 4636 4645 4654 4663 4672 4681 4690 4700 4709 4718 4727 4736 4745 4754 4763 4772 4781 4790 4800 4809 4818 4827 4836 4845 4854 4863 4872 4881 4890 4900 4909 4918 4927 4936 4945 4954 4963 4972 4981 4990 5000 5009 5018 5027 5036 5045 5054 5063 5072 5081 5090 5100 5109 5118 5127 5136 5145 5154 5163 5172 5181 5190 5200 5209 5218 5227 5236 5245 5254 5263 5272 5281 5290 5300 5309 5318 5327 5336 5345 5354 5363 5372 5381 5390 5400 5409 5418 5427 5436 5445 5454 5463 5472 5481 5490 5500 5509 5518 5527 5536 5545 5554 5563 5572 5581 5590 5600 5609 5618 5627 5636 5645 5654 5663 5672 5681 5690 5700 5709 5718 5727 5736 5745 5754 5763 5772 5781 5790 5800 5809 5818 5827 5836 5845 5854 5863 5872 5881 5890 5900 5909 5918 5927 5936 5945 5954 5963 5972 5981 5990 6000 6009 6018 6027 6036 6045 6054 6063 6072 6081 6090 6100 6109 6118 6127 6136 6145 6154 6163 6172 6181 6190 6200 6209 6218 6227 6236 6245 6254 6263 6272 6281 6290 6300 6309 6318 6327 6336 6345 6354 6363 6372 6381 6390 6400 6409 6418 6427 6436 6445 6454 6463 6472 6481 6490 6500 6509 6518 6527 6536 6545 6554 6563 6572 6581 6590 6600 6609 6618 6627 6636 6645 6654 6663 6672 6681 6690 6700 6709 6718 6727 6736 6745 6754 6763 6772 6781 6790 6800 6809 6818 6827 6836 6845 6854 6863 6872 6881 6890 6900 6909 6918 6927 6936 6945 6954 6963 6972 6981 6990 7000 7009 7018 7027 7036 7045 7054 7063 7072 7081 7090 7100 7109 7118 7127 7136 7145 7154 7163 7172 7181 7190 7200 7209 7218 7227 7236 7245 7254 7263 7272 7281 7290 7300 7309 7318 7327 7336 7345 7354 7363 7372 7381 7390 7400 7409 7418 7427 7436 7445 7454 7463 7472 7481 7490 7500 7509 7518 7527 7536 7545 7554 7563 7572 7581 7590 7600 7609 7618 7627 7636 7645 7654 7663 7672 7681 7690 7700 7709 7718 7727 7736 7745 7754 7763 7772 7781 7790 7800 7809 7818 7827 7836 7845 7854 7863 7872 7881 7890 7900 7909 7918 7927 7936 7945 7954 7963 7972 7981 7990 8000 8009 8018 8027 8036 8045 8054 8063 8072 8081 8090 8100 8109 8118 8127 8136 8145 8154 8163 8172 8181 8190 8200 8209 8218 8227 8236 8245 8254 8263 8272 8281 8290 8300 8309 8318 8327 8336 8345 8354 8363 8372 8381 8390 8400 8409 8418 8427 8436 8445 8454 8463 8472 8481 8490 8500 8509 8518 8527 8536 8545 8554 8563 8572 8581 8590 8600 8609 8618 8627 8636 8645 8654 8663 8672 8681 8690 8700 8709 8718 8727 8736 8745 8754 8763 8772 8781 8790 8800 8809 8818 8827 8836 8845 8854 8863 8872 8881 8890 8900 8909 8918 8927 8936 8945 8954 8963 8972 8981 8990 9000 9009 9018 9027 9036 9045 9054 9063 9072 9081 9090 9100 9109 9118 9127 9136 9145 9154 9163 9172 9181 9190 9200 9209 9218 9227 9236 9245 9254 9263 9272 9281 9290 9300 9309 9318 9327 9336 9345 9354 9363 9372 9381 9390 9400 9409 9418 9427 9436 9445 9454 9463 9472 9481 9490 9500 9509 9518 9527 9536 9545 9554 9563 9572 9581 9590 9600 9609 9618 9627 9636 9645 9654 9663 9672 9681 9690 9700 9709 9718 9727 9736 9745 9754 9763 9772 9781 9790 9800 9809 9818 9827 9836 9845 9854 9863 9872 9881 9890 9900 9909 9918 9927 9936 9945 9954 9963 9972 9981 9990 10000 10009 10018 10027 10036 10045 10054 10063 10072 10081 10090 10100 10109 10118 10127 10136 10145 10154 10163 10172 10181 10190 10200 10209 10218 10227 10236 10245 10254 10263 10272 10281 10290 10300 10309 10318 10327 10336 10345 10354 10363 10372 10381 10390 10400 10409 10418 10427 10436 10445 10454 10463 10472 10481 10490 10500 10509 10518 10527 10536 10545 10554 10563 10572 10581 10590 10600 10609 10618 10627 10636 10645 10654 10663 10672 10681 10690 10700 10709 10718 10727 10736 10745 10754 10763 10772 10781 10790 10800 10809 10818 10827 10836 10845 10854 10863 10872 10881 10890 10900 10909 10918 10927 10936 10945 10954 10963 10972 10981 10990 11000 11009 11018 11027 11036 11045 11054 11063 11072 11081 11090 11100 11109 11118 11127 11136 11145 11154 11163 11172 11181 11190 11200 11209 11218 11227 11236 11245 11254 11263 11272 11281 11290 11300 11309 11318 11327 11336 11345 11354 11363 11372 11381 11390 11400 11409 11418 11427 11436 11445 11454 11463 11472 11481 11490 11500 11509 11518 11527 11536 11545 11554 11563 11572 11581 11590 11600 11609 11618 11627 11636 11645 11654 11663 11672 11681 11690 11700 11709 11718 11727 11736 11745 11754 11763 11772 11781 11790 11800 11809 11818 11827 11836 11845 11854 11863 11872 11881 11890 11900 11909 11918 11927 11936 11945 11954 11963 11972 11981 11990 12000 12009 12018 12027 12036 12045 12054 12063 12072 12081 12090 12100 12109 12118 12127 12136 12145 12154 12163 12172 12181 12190 12200 12209 12218 12227 12236 12245 12254 12263 12272 12281 12290 12300 12309 12318 12327 12336 12345 12354 12363 12372 12381 12390 12400 12409 12418 12427 12436 12445 12454 12463 12472 12481 12490 12500 12509 12518 12527 12536 12545 12554 12563 12572 12581 12590 12600 12609 12618 12627 12636 12645 12654 12663 12672 12681 12690 12700 12709 12718 12727 12736 12745 12754 12763 12772 12781 12790 12800 12809 12818 12827 12836 12845 12854 12863 12872 12881 12890 12900 12909 12918 12927 12936 12945 12954 12963 12972 12981 12990 13000 13009 13018 13027 13036 13045 13054 13063 13072 13081 13090 13100 13109 13118 13127 13136 13145 13154 13163 13172 13181 13190 13200 13209 13218 13227 13236 13245 13254 13263 13272 13281 13290 13300 13309 13318 13327 13336 13345 13354 13363 13372 13381 13390 13400 13409 13418 13427 13436 13445 13454 13463 13472 13481 13490 13500 13509 13518 13527 13536 13545 13554 13563 13572 13581 13590 13600 13609 13618 13627 13636 13645 13654 13663 13672 13681 13690 13700 13709 13718 13727 13736 13745 13754 13763 13772 13781 13790 13800 13809 13818 13827 13836 13845 13854 13863 13872 13881 13890 13900 13909 13918 13927 13936 13945 13954 13963 13972 13981 13990 14000 14009 14018 14027 14036 14045 14054 14063 14072 14081 14090 14100 14109 14118 14127 14136 14145 14154 14163 14172 14181 14190 14200 14209 14218 14227 14236 14245 14254 14263 14272 14281 14290 14300 14309 14318 14327 14336 14345 14354 14363 14372 14381 14390 14400 14409 14418 14427 14436 14445 14454 14463 14472 14481 14490 14500 14509 14518 14527 14536 14545 14554 14563 14572 14581 14590 14600 14609 14618 14627 14636 14645 14654 14663 14672 14681 14690 14700 14709 14718 14727 14736 14745 14754 14763 14772 14781 14790 14800 14809 14818 14827 14836 14845 14854 14863 14872 14881 14890 14900 14909 14918 14927 14936 14945 14954 14963 14972 14981 14990 15000 15009 15018 15027 15036 15045 15054 15063 15072 15081 15090 15100 15109 15118 15127 15136 15145 15154 15163 15172 15181 15190 15200 15209 15218 15227 15236 15245 15254 15263 15272 15281 15290 15300 15309 15318 15327 15336 15345 15354 15363 15372 15381 15390 15400 15409 15418 15427 15436 15445 15454 15463 15472 15481 15490 15500 15509 15518 15527 15536 15545 15554 15563 15572 15581 15590 15600 15609 15618 15627 15636 15645 15654 15663 15672 15681 15690 15700 15709 15718 15727 15736 15745 15754 15763 15772 15781 15790 15800 15809 15818 15827 15836 15845 15854 15863 15872 15881 15890 15900 15909 15918 15927 15936 15945 15954 15963 15972 15981 15990 16000 16009 16018 16027 16036 16045 16054 16063 16072 16081 16090 16100 16109 16118 16127 16136 16145 16154 16163 16172 16181 16190 16200 16209 16218 16227 16236 16245 16254 16263 16272 16281 16290 16300 16309 16318 16327 16336 16345 16354 16363 16372 16381 16390 16400 16409 16418 16427 16436 16445 16454 16463 16472 16481 16490 16500 16509 16518 16527 16536 16545 16554 16563 16572 16581 16590 16600 16609 16618 16627 16636 16645 16654 16663 16672 16681 16690 16700 16709 16718 16727 16736 16745 16754 16763 16772 16781 16790 16800 16809 16818 16827 16836 16845 16854 16863 16872 16881 16890 16900 16909 16918 16927 16936 16945 16954 16963 16972 16981 16990 17000 17009 17018 17027 17036 17045 17054 17063 17072 17081 17090 17100 17109 17118 17127 17136 17145 17154 17163 17172 17181 17190 17200 17209 17218 17227 17236 17245 17254 17263 17272 17281 17290 17300 17309 17318 17327 17336 17345 17354 17363 17372 17381 17390 17400 17409 17418 17427 17436 17445 17454 17463 17472 17481 17490 17500 17509 17518 17527 17536 17545 17554 17563 17572 17581 17590 17600 17609 17618 17627 17636 17645 17654 17663 17672 17681 17690 17700 17709 17718 17727 17736 17745 17754 17763 17772 17781 17790 17800 17809 17818 17827 17836 17845 17854 17863 17872 17881 17890 17900 17909 17918 17927 17936 17945 17954 17963 17972 17981 17990 18000 18009 18018 18027 18036 18045 18054 18063 18072 18081 18090 18100 18109 18118 18127 18136 18145 18154 18163 18172 18181 18190 18200 18209 18218 18227 18236 18245 18254 18263 18272 18281 18290 18300 18309 18318 18327 18336 18345 18354 18363 18372 18381 18390 18400 18409 18418 18427 18436 18445 18454 18463 18472 18481 18490 18500 18509 18518 18527 18536 18545 18554 18563 18572 18581 18590 18600 18609 18618 18627 18636 18645 18654 18663 18672 18681 18690 18700 18709 18718 18727 18736 18745 18754 18763 18772 18781 18790 18800 18809 18818 18827 18836 18845 18854 18863 18872 18881 18890 18900 18909 18918 18927 18936 18945 18954 18963 18972 18981 18990 19000 19009 19018 19027 19036 19045 19054 19063 19072 19081 19090 19100 19109 19118 19127 19136 19145 19154 19163 19172 19181 19190 19200 19209 19218 19227 19236 19245 19254 19263 19272 19281 19290 19300 19309 19318 19327 19336 19345 19354 19363 19372 19381 19390 19400 19409 19418 19427 19436 19445 19454 19463 19472 19481 19490 19500 19509 19518 19527 19536 19545 19554 19563 19572 19581 19590 19600 19609 19618 19627 19636 19645 19654 19663 19672 19681 19690 19700 19709 19718 19727 19736 19745 19754 19763 19772 19781 19790 19800 19809 19818 19827 19836 19845 19854 19863 19872 19881 19890 19900 19909 19918 19927 19936 19945 19954 19963 19972 19981 19990 20000 20009 20018 20027 20036 20045 20054 20063 20072 20081 20090 20100 20109 20118 20127 20136 20145 20154 20163 20172 20181 20190 20200 20209 20218 20227 20236 20245 20254 20263 20272 20281 20290 20300 20309 20318 20327 20336 20345 20354 20363 20372 20381 20390 20400 20409 20418 20427 20436 20445 20454 20463 20472 20481 20490 20500 20509 20518 20527 20536 20545 20554 20563 20572 20581 20590 20600 20609 20618 20627 20636 20645 20654 20663 20672 20681 20690 20700 20709 20718 20727 20736 20745 20754 20763 20772 20781 20790 20800 20809 20818 20827 20836 20845 20854 20863 20872 20881 20890 20900 20909 20918 20927 20936 20945 20954 20963 20972 20981 20990 21000 21009 21018 21027 21036 21045 21054 21063 21072 21081 21090 21100 21109 21118 21127 21136 21145 21154 21163 21172 21181 21190 21200 21209 21218 21227 21236 21245 
```


RT "Single-chain structure of human ceruloplasmin: the complete amino
 RT acid sequence of the whole molecule."; [6]
 RN Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
 RP SEQUENCE OF 158-333; 518-724 AND 858-1065.
 RX MEDLINE=83117800; PubMed=6571985;
 RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
 RA Putnam F.W.;
 RA "Internal triplication in the structure of human ceruloplasmin."; [7]
 RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
 RN SEQUENCE OF 501-905.
 RP MEDLINE=81199407; PubMed=6940148;
 RX Dwulet F.E., Putnam F.W.;
 RA "Complete amino acid sequence of a 50,000-dalton fragment of human
 RT ceruloplasmin."; [8]
 RN Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
 RP SEQUENCE OF 907-1065.
 RX MEDLINE=80137543; PubMed=6987229;
 RA Kingston I.B., Kingston B.L., Putnam F.W.;
 RA "Primary structure of a histidine-rich proteolytic fragment of human
 RT ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
 RT peptides."; [9]
 RL J. Biol. Chem. 255:2878-2885(1980).
 RN SEQUENCE OF 1007-1061 FROM N.A.
 RP MEDLINE=90285218; PubMed=2355023;
 RX Yang F.W., Friedrichs W.E., Cupples R.L., Banifacio M.J.,
 RA Sanford J.A., Horton W.A., Bowman B.H.;
 RA "Human ceruloplasmin. Tissue-specific expression of transcripts
 RT produced by alternative splicing."; [10]
 RL J. Biol. Chem. 265:10780-10785(1990).
 RN REVIEW.
 RP MEDLINE=22049919; PubMed=12055353;
 RX Hellman N.E., Gitlin J.D.;
 RA "Ceruloplasmin metabolism and function."; [12]
 RL Annu. Rev. Nutr. 22:439-458(2002).
 RN X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
 RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
 RA Lindley P.;
 RA "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of
 RT the copper centres."; [13]
 RL J. Biol. Inorg. Chem. 1:15-23(1996).
 CC -!- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
 CC MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
 CC ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
 CC AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
 CC -!- COPACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
 CC THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS
 CC KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPERS
 CC BINUCLEAR.
 CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
 CC PLASMA.
 CC -!- DISEASE: Defects in CP are the cause of aceruloplasminemia
 CC [MIM:604290], an autosomal recessive disorder of iron metabolism.
 CC It is characterized by iron accumulation in the brain as well as
 CC visceral organs. Clinical features consist of the triad of retinal
 CC degeneration, diabetes mellitus and neurological disturbances.
 CC -!- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,
 CC in which copper cannot be incorporated into ceruloplasmin in
 CC liver because of defects in the copper-transporting ATPase 2.
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M13699; AAA51976.1; -
 DR EMBL; D45045; BAA08085.1; -
 DR EMBL; D45044; BAA08084.1; -
 DR EMBL; D45028; BAA08084.1; JOINED.
 DR EMBL; D45029; BAA08084.1; JOINED.
 DR EMBL; D45030; BAA08084.1; JOINED.
 DR EMBL; D45031; BAA08084.1; JOINED.
 DR EMBL; D45032; BAA08084.1; JOINED.
 DR EMBL; D45033; BAA08084.1; JOINED.
 DR EMBL; D45034; BAA08084.1; JOINED.
 DR EMBL; D45035; BAA08084.1; JOINED.
 DR EMBL; D45036; BAA08084.1; JOINED.
 DR EMBL; D45037; BAA08084.1; JOINED.
 DR EMBL; D45038; BAA08084.1; JOINED.
 DR EMBL; D45039; BAA08084.1; JOINED.
 DR EMBL; D45040; BAA08084.1; JOINED.
 DR EMBL; D45041; BAA08084.1; JOINED.
 DR EMBL; D45042; BAA08084.1; JOINED.
 DR EMBL; D45043; BAA08084.1; JOINED.
 DR EMBL; D00025; BAA00019.1; -
 DR EMBL; X04135; CAA27752.1; -
 DR EMBL; X04136; CAA27753.1; -
 DR EMBL; X04137; CAA27754.1; -
 DR EMBL; X04138; CAA27755.1; -
 DR EMBL; M13536; AAA51975.1; -
 DR PIR; A25443; K0HU.
 DR PDB; 1KCW; 12-FEB-97.
 DR GLYCSuiteDB; P00450; -
 DR SWISS-2DPAGE; P00450; HUMAN.
 DR Siena-2DPAGE; P00450; -
 DR Genew; HGNC:2295; CP.
 DR MIM; 117700; -
 DR MIM; 604290; -
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0006879; P:iron ion homeostasis; TAS.
 DR InterPro; IPR001117; Cu-oxidase.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 DR Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;
 KW Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 1065 CERULOPLASMIN.
 FT F5/8 TYPE A 1.
 FT DOMAIN 20 357 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 20 397 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 209 357
 FT DOMAIN 370 718 F5/8 TYPE A 2.
 FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 730 1061 F5/8 TYPE A 3.
 FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .).
 FT DISULFID 174 200 PROBABLE.
 FT DISULFID 276 357 PROBABLE.
 FT DISULFID 534 560 PROBABLE.
 FT DISULFID 637 718 PROBABLE.
 FT DISULFID 874 900 PROBABLE.
 FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).

AC Q61147;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
 GN CP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Klomp L.W.J., Fathangrazi Z.S., Choi D.W., Gitlin J.D.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=96294736; PubMed=8690795;
 RA Klomp L.W.J., Fathangrazi Z.S., Dugan L.L., Gitlin J.D.;
 RT "Ceruloplasmin gene expression in the murine central nervous system.";
 RL J. Clin. Invest. 98:207-215(1996).
 CC -!- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
 MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
 ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
 AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
 CC -!- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
 THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS
 KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED
 BINUCLEAR.
 CC -!- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.
 CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U49430; AAB07996.1; .
 DR HSP; P00450; 1KCW.
 DR MGD; MGI:89476; Cn.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu_oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 KW Oxidoreductase; Copper; Metal-binding; Glycoprotein; Plasma; Repeat;
 KW Signal.
 FT SIGNAL. 1 19 BY SIMILARITY.
 FT CHAIN 20 1062 CERULOPLASMIN.
 FT DOMAIN 20 356 F5/8 TYPE A 1.
 FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 208 356 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 369 713 F5/8 TYPE A 2.
 FT DOMAIN 369 555 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 565 713 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 725 1057 F5/8 TYPE A 3.
 FT DOMAIN 725 896 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 904 1057 PLASTOCYANIN-LIKE 6.
 FT DISULFID 173 199 BY SIMILARITY.
 FT DISULFID 275 356 BY SIMILARITY.
 FT DISULFID 529 555 BY SIMILARITY.
 FT DISULFID 632 713 BY SIMILARITY.
 FT DISULFID 870 896 BY SIMILARITY.
 FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 179 179 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 990 990 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 993 993 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 995 995 COPPER (TYPE 3) (BY SIMILARITY).

FT	METAL	1035	1035	COPPER (TYPE 3) (BY SIMILARITY).
FT	METAL	1036	1036	COPPER (TYPE 1) (BY SIMILARITY).
FT	METAL	1037	1037	COPPER (TYPE 3) (BY SIMILARITY).
FT	METAL	1041	1041	COPPER (TYPE 1) (BY SIMILARITY).
FT	METAL	1046	1046	COPPER (TYPE 1) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	226	226	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	583	583	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	625	625	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	757	757	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	922	922	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1062	AA; 121159 MW; F3F52ED09A238F16 CRC64;	

Query Match 21.3%; Score 1635.5; DB 1; Length 1062;
 Best Local Similarity 32.4%; Pred. No. 4.8e-99;
 Matches 376; Conservative 195; Mismatches 438; Indels 153; Gaps 24;

QY	3	RYRILGAVELSDVMQ--SDLGELPUDARPPRPVKSPFNTSVVYKTLFVEVTHLEN	60
Db	22	KHPFGITEAVMDYASCTEEKKLSVDTQSNFYLQNGPDRIGRKYKALYFEYTDGTF	81
QY	61	IAKPRPPMGLIGPTIOAEVVDVTWITLKNWASHPSLHAGVSYMKASEAGAEYDDOTSQ	120
Db	82	KTIIDPAWGLGVLGVIKAEVEDKVVHLKNLASIYTFHAGVITYYKEYEGANVPDNTD	141
QY	121	REKEDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNGLIGALLVCR	180
Db	142	FORADDDKVLFGQYVYVHLA--NEPSGEGDSNCVTRIYHSHVDAPAKDIASGLPLILCK	200
QY	181	EGSLAKETQTL--HKPILLFAVEDEGKSHSETKSLM-----QDRQASARAWPMGHT	233
Db	201	KGSLYKKEKKNIDQEFVLMFVSVDENLSWYLENITFCSEPEKVDKDNEDFQESNRMYS	260
QY	234	VNGVYVNSLPGLICHRKSVVHYVIGMTTPEVHSIFLEGHTFLVRNHRQASLSIPTF	293
Db	261	INGYTFGLSGLSCAADRVKVKYLFNGNEVDVHSAFFHGQALTSRYNQTDIINLPATL	320
QY	294	LTAQTLMDLGQFLFCHISHQHDGMEAYVKVDSCEEPQLRMKNNEABDYDDDLDS	353
Db	321	IDATVVAQNFQVWMLSCQNLNHLKAGLQAFQVQDC-----NKPESKD-----	363
QY	354	EMDVVRFDNDDNSPSFIOIRSVAKKHPTWVHYIAAEEDMDYAP--LVLAAPDDR-----S	406
Db	364	-----NIRGHVRH-----YYIAAEVINYIAPSGIDIFTEELKTASGS	402
QY	407	YKSOYLNNGFQIRIGRKYKRVFMAVYDTDF---KTREAIQHSIGILGFLYGEVDTLLI	463
Db	403	DSGVFPEQGATRIGGSYKVMAYREYTDGSGFTNRKERGPDEHGLGILGPVIAEVGDIKV	462
QY	464	IFKQASRPVNIYPHGIT---DVRPLYSRRLPGVHKLKDP--ILPGEIFKYKWTVTVE	518
Db	463	TFHNKGQHLSIQPMGVSFPTAENEGTYG---PFGASSQQAASHVAPKXTFYEWETPK	519
QY	519	DGTPKSDPRCLTRYSSFVNMERDASGLIGLILCYKESVDQSGNOIMSDKRVILPSV	578
Db	520	MGPTIADPVLCKMYSAVDPDKDIFGLIGPKLCKGSLADGRQKVDKDYELFPFV	579
QY	579	FDEHRSWYLATENIORPLPNPAGVQLEDPFQASNMHSINGYVFDLSQL--SVCLHREWAYV	637
Db	580	FDENESLLDDNIRMFTAPDQVDKEDDFQESNKHSHMNGFMVGNQSWHMCMLGESIVW	639
QY	638	YILSIGAQTDPLSVFSGTYFKKGVYVEDTLTFPFSGETVFMSEMPGLWILGCHNSDP	697
Db	640	YLFSGAGNEADVHGLYFSGNTYLCCKGEERDTANLFFPKSLTLLMNPDKTGTFDVECLT	699
QY	698	RNRGWTALLKVSCKDKTGDYEDSDISAYLLSKNNAIEPRSFSONPVLKRQREIT	757
Db	700	YTGKMKQKTYVNCQR-----QFEDFTVYGER-----	727
QY	758	RTTLQSQOEIDYDDTISVEMKKEDFDIYDEDNQSPRSQKTRHYPIAAVERLMDYGM	817
Db	728	--TYVDVAVEWMDYSPSRAWEKELHLL--QEQVNSNVFLOKE--EFFIGS-----	772

QY 818 SSSPHVLRNRAQSGVPOFKVVFQFETDGSFTQPL- YRGELNEHLGLGPVIRAEVDN 876
 DB 773 -----KIKVVYHQFTDSFREQVKRAEDEDHLGILGPPPIHANVGDK 815
 QY 877 IMVTFRNOASRPSPYSSLSIYEEDQOAEPRKNFV---KPNETKTYFWKVQHMMAPTK 933
 DB 816 VKVWFKNMTRPYSIHA-----HGVTKESSTVVTLPGEVATYTWQIPERSGAGR 865
 QY 934 DEFQCKAYESDVLEKDVHSLGLIGLLVC---HTNTLNPAGHQVTVQEPALFTTFD 990
 DB 866 EDSACIPWAYSTVDVNDLYSLGLPLIVCKRSYVKVFSK-----KMEFFLLFLVFD 920
 QY 991 ETKSWYFENMERNCRAPCNQMEDPTFKENYRFAINGYIMDTLPGLVMAQDQIRWYL 1050
 DB 921 ENESYLDNNTKITYSEHEKYNKDEEFLESKMEHANGKMFNQLGLTMHVKDEVNWL 980
 QY 1051 LSMGSNEHSHFSGHVTFRKBEYKQALYNLPGVFTVEMLPKAGIWRVECLIGE 1110
 DB 981 MGNMNEIDLHTVHFRGHSGFYKHGRGVYSSDVPDLFPFGTYQTLEMPFQTPGTTWLLHCHVTD 1040
 QY 1111 HLHAGMSTLFLVYSNKCOTPLG 1132
 DB 1041 HVHAGMATTYVLPVEQETKSG 1062

RESULT 10
 MFGM_RAT
 ID MFGM_RAT STANDARD; PRT; 427 AA.
 AC P70490;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-acetyl GD3 ganglioside synthase) (AGS) (MFGM).
 GN MFG88 OR AGS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96374422; PubMed=8780713;
 RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
 RT "Cloning and expression of cDNA for O-acetylation of GD3 ganglioside";
 RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
 CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.

 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; D84068; BAAL2210.1; -.
 CC F01; JC4915; JC4915.
 CC HSSP; P00740; 1EDM.
 CC InterPro; IPR006209; EGF like.
 CC InterPro; IPR000421; FAS8 C.
 CC InterPro; IPR006210; IEGF.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00754; F5_F8 type C; 2.
 CC SMART; SM00181; EGF; 2.
 CC SMART; SM00231; FAS8C; 2.
 CC PROSITE; P500022; EGF_1; 2.

DR PROSITE; PS01186; EGF 2; 2.
 DR PROSITE; PS01285; FAS8C_1; 2.
 DR PROSITE; PS01286; FAS8C_2; 2.
 DR PROSITE; PS50022; FAS8C_3; 2.
 KW SIGNAL; Glycoprotein; Repeat; EGF-like domain; Milk. POTENTIAL.
 FT SIGNAL 1 22
 FT CHAIN 23 427
 FT DOMAIN 24 61
 FT DOMAIN 64 108
 FT DOMAIN 111 267
 FT DOMAIN 272 427
 FT DISULFID 28 39
 FT DISULFID 33 49
 FT DISULFID 51 60
 FT DISULFID 68 79
 FT DISULFID 73 96
 FT DISULFID 98 107
 FT DISULFID 111 267
 FT DISULFID 254 258
 FT DISULFID 272 427
 FT SITE 87 89
 FT CARBOHYD 61 61
 FT CARBOHYD 230 230
 FT CARBOHYD 280 280
 FT CARBOHYD 390 390
 SQ SEQUENCE 427 AA; 47413 MW; EA8C9631F3EB6047 CRC64;

Query Match 8.6%; Score 663; DB 1; Length 427;
 Best Local Similarity 40.3%; Pred. No. 5.5e-36;
 Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
 QY 1105 ECLGEHLHAGMSTLFLVYSNK-----CQPLGWSGHIRDFOITASGOY-- 1149
 DB 78 KCLVETDQSG--DIFTEYICQCPGVSGIHCELCSTKLGEGAIADSISSSVVMG 135
 QY 1150 ---GQMAPKLARLHYSGINAW--STKEPFSWIKVDLLAPMIHGIKTQGARKFSSLY 1203
 DB 136 FWGLQRMGPELRLVYRTGIVNAVNTASSYSDSKPQIOWFLRKMVSGVMTQGASRAGAEY 195
 QY 1204 ISQFTIMYSLDGKKQWYRGNSTGTLMVYFPGNVDSGIGIKNIENPPPIARVIRLPHPTYS 1263
 DB 196 LKTFKWAYSILDRGRPEFIQDESQDGEFGMGNQNNLSKINFNPTLEAQVIRLPPVSC 255
 QY 1264 IRSTRLMELMGCDLNSCSMPLMBSKATSDAQITASSYFT--NMFA--TWSPSKARHLQ 1320
 DB 256 RGCTLRPELLGCELCGCEPLGLKNNTIPDSQITASSYKTNLRAFGWYHLGLDNOG 315
 QY 1321 RSNAWRPQVNNPKWLQVDFQTKMKTGVTGTTQGVKSLTSMYVKEFLSSQDGHQWTLF 1380
 DB 316 KINAWTAQSNKAKWLQVLDLGTQKKVTGIIITQGARDFGHIQYVASYKVVAHSDGQWTVY 375
 QY 1381 PQNGKVKVFGNQDSFTFVNVSLDPLLRVLRHPSQSWVHQIALRMEVLGC 1432
 DB 376 EEQGTSKVFGQNLNNSHKKNIFEKPFVARYVRVLPVLRVLPVLRVLRVLRVLRVLRV 427

RESULT 11
 MFGM_MOUSE
 ID MFGM_MOUSE STANDARD; PRT; 463 AA.
 AC P21956; P97800;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
 DE (Sperm surface protein SP47) (MP47).
 GN MFG88.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
 RC TISSUE=Mammary gland;

```
RX MEDLINE=91046008; PubMed2122462;
RA Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
RT reveals the existence of epidermal growth factor-like domains linked
RT to factor VIII-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN [2]
RN SEQUENCE OF 23-463 FROM N.A.
RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38337; AAA39534.1; --
DR EMBL; Y11684; CAA72380.1; --
DR PIR; A36479; A36479.
DR HSP; P00740; 1EDM.
DR MGD; MGI:102768; Mfge8.
DR InterPro; IPR001438; EGF II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000421; FAS5 C.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 2.
DR PRINTS; PF00754; F5_F8 type C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS5C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FAS5C_1; 2.
DR PROSITE; PS01286; FAS5C_2; 2.
DR PROSITE; PS00022; FAS5C_3; 2.
DR PROSITE; PS00022; FAS5C_3; 2.
DR Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22
FT CHAIN 23 463 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT DOMAIN 148 303 F5/8 TYPE C 1.
FT DOMAIN 308 463 F5/8 TYPE C 2.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 148 303 BY SIMILARITY.
FT DISULFID 290 424 BY SIMILARITY.
FT DISULFID 308 463 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 30 30 S -> F (IN REF. 2).
FT CONFLICT 35 35 N -> D (IN REF. 1; AA SEQUENCE).
FT CONFLICT 110 147 ETVNYLDGEYMETTAVPTAVPTAPDLSNNLASR ->
-----
FT CONFLICT 168 168 G (IN REF. 2).
FT CONFLICT 196 196 Y -> S (IN REF. 2).
FT CONFLICT 309 309 H -> T (IN REF. 2).
FT CONFLICT 395 395 L -> S (IN REF. 2).
FT CONFLICT 463 463 E -> A (IN REF. 2).
SQ SEQUENCE 463 AA; 51465 MW; D78B6C6EFBBA724D CRC64;
-----
Query Match 8.5%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 1.5e-35;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;
-----
QY 1124 SNKCTPLGMAAGCHTRDFOQTASGOY-----GQWAPKLRLHYSGSNAM--STKEPFS 1175
DB 145 ASRCSTQLGMEGAIADSIASIVYMGFMGLQRMGLARLYRTGIYNHASNYSK 204
QY 1176 MKVLLAPMIITHGKTQARCFESLYISPIINYSIDGKKQVVRGNTGTLVAFGN 1235
DB 205 WQVNLRLKRVSGVWVTOGASRAGAEVLTQFKVAYSIDGRKFEIQDESGD-KEFLGN 263
QY 1236 VDSSGINKINFPPIIYIRLHPHYISIRSTLRMLMGCDLNSCMPLGMSKASDAQ 1295
DB 264 LDNNSLKVMNRPTEAQYIRLYPVSCRGCTLRPELLGCEHLGCLPLGLKNNITPDSQ 323
QY 1296 ITASSYFT--NWFA-TWSPSKARLHLOQRSHNRQVNPVNPKEWLOVDQKTMKVTGVTQ 1352
DB 324 WSASSSYKTNLRAFQWYFHLGRDNLNQKINAWTAQSNKAEWLQVDLGTQRTQVGTITQ 383
QY 1353 GVKSLTSMYKFEFLISSQGHQHTLFFQNGKVKVFGQGNODSFTFVNSLDPLLLRYL 1412
DB 384 GARDGCHQYVESYKVAHSDGQVTVYERQSSKVFQGNLDNNSHKNIFEKPMARYV 443
QY 1413 RIHQSWVHQIALRMEVLGC 1432
DB 444 RVLPSVWNRITRLRELLGC 463
-----
RESULT 12
MFGM_PIG
ID MFGM_PIG STANDARD; PRT; 409 AA.
AC P79385;
DT 01-JUN-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin (Milk fat Globule-EGF factor 8) (MFG-8) (MFGM) (Sperm
DE surface protein SP47) (PP47).
GN MFG8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11683; CAA72379.1; --
DR PIR; T11743; T11743.
```

```

DR HSP; P00740; 1EDM.
DR InterPro: IPR006209; EGF like.
DR InterPro: IPR000421; FA5E C.
DR InterPro: IPR006210; IEGF.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 88 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; B0C07AF80029927A CRC64;

Query Match 8.5%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 3.7e-35;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 1105 ECLIGHLHAGSTLFLVYSNK-----CQTELGWASGHIRDFQITAGQY--- 1149
Db 58 ELEVDDAARG--DVFTYIICKPHCYTGTHCEILCNAPLGWETGAIDFQASSMHIG 115
QY 1150 ----GQWAPKLARLHYSGINAW--STKEPFFSWIKVDLAPMIIRGKTKQARQKFSSLY 1203
Db 116 FMGLQWAPELARLHRAGIVNAWTASNYDRNPWQVNLRRMRVTGVVVTQGRASRAGSY 175
QY 1204 ISQFIWYSLDKKQTYGNTGTLWFFGNVDSGKHNFPPIIARYIRLPHYTHYS 1263
Db 176 MKTFKVAYSTDKRFQFQIGAEESGDKIFMGLDNLGSLKVNLFEPVLEVQYVRLVPIICH 235
QY 1264 IRSLRMELMGCDLNSCPLNGESKAIISDAOITASSYFTN--MEATWSPSKARHLQX 1320
Db 236 RGCTLRPELGLGELSCAEPLGKONTIPNKQITASSFRTWGLSAPSWYFYALDNDQ 295
QY 1321 RSNAMRPQNNPKMLQVDFQTKMKTGVTTQGVKSLTSMYVKEFLISSODGHQWTLF 1380
Db 296 KFNWTAQNSASEMLQIDLGSRQRTVGIITQGRDFGHIQVAAKVAYSDDGVSWTEY 355
QY 1381 FQNGKV--KVFQGNODSFPVNVNSLDPPLTYLRIHQSWHQAALRNEVLGC 1432
Db 356 RDQGALEGIKFFGNLDNNSHKMPELTPTRFVRLPVAMHNRITLRAVELLGC 409

RESULT 13
MFGM_BOVIN
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; P79344; Q27959;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MGF57/53) (PAS-6/PAS-7 Glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.

```

```

OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN-Holstein; TISSUE-Mammary gland;
RA MEDLINE=97008954; PubMed=8856064;
RX Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE-Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Ackl N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE-Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGP-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms-2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X91895; CAA62997.1; -.
DR EMBL; S80643; AAB35894.2; -.
DR EMBL; Y11719; CAA72406.1; -.
DR PIR; S74211; S74211.
DR HSP; P00740; 11XA.
DR InterPro: IPR006209; EGF like.
DR InterPro: IPR000421; FA58 C.
DR InterPro: IPR006210; IEGF.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.

```



```
DR PROSITE; PS01286; FA58C 2; 2.  
DR PROSITE; PS50022; FA58C 3; 2.  
KW Signal: Glycoprotein; Milk; Repeat; EGF-like domain;  
KW Alternative splicing.  
FT SIGNAL 1 18  
FT CHAIN 19 427 LACTADHERIN.  
FT DOMAIN 20 59 EGF-LIKE 1.  
FT DOMAIN 62 106 EGF-LIKE 2.  
FT DOMAIN 109 265 F5/8 TYPE C 1.  
FT DOMAIN 270 427 F5/8 TYPE C 2.  
FT SITE 85 97 CELL ATTACHMENT SITE (POTENTIAL).  
FT DISULFID 24 35 BY SIMILARITY.  
FT DISULFID 29 47 BY SIMILARITY.  
FT DISULFID 49 58 BY SIMILARITY.  
FT DISULFID 66 77 BY SIMILARITY.  
FT DISULFID 71 94 BY SIMILARITY.  
FT DISULFID 96 105 BY SIMILARITY.  
FT DISULFID 109 265 BY SIMILARITY.  
FT DISULFID 252 256  
FT DISULFID 270 427  
FT CARBOHYD 27 27  
FT CARBOHYD 34 34  
FT CARBOHYD 59 59  
FT CARBOHYD 227 227  
FT VARSPLIC 169 221  
FT CONFLICT 19 19  
FT CONFLICT 28 28  
FT SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4EB24 CRC64;  
Query Match 8.3%; Score 635; DB 1; Length 427;  
Best Local Similarity 37.9%; Pred. No. 3.8e-34;  
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;  
QY 1105 ECLIGHLHAGSTFLVYSNK-----CQPLGHASGHIRDFQITASQY-- 1149  
DB 76 EQVTDSDSHRG--DVPIQYICKPLGVGVGHCHCTTCTPLGMQTGAIDQSASSMHIG 133  
QY 1150 ---GQWAPKLARLHYSGSINAWST---KEPPSWIKVDLLAPMIHGKTQCARQKFS 1201  
DB 134 FWGLQWAEARLALHGTGIVNATWSGNYDKNP--NIOVLMKRWGTGVVTOGSEAGSA 191  
QY 1202 LYISQFTIMYSLDGKKQWYRGNSGTGLMVFGVNDSSGIGKNIENPPIIARYIRLHPH 1261  
DB 192 EYLTKEFVAYTDGROFQIQVAGRSQDKIFIGNVNSGLKINLEFPLETQYVRLVPII 251  
QY 1262 YSIRSLRMLMGLDNLSCMPLGMSKASDAQITASSYFTN---MPATWSPSKARLHL 1318  
DB 252 CHRGCTLRFLGCELGNGCTEPGLKONTIPKQITASSYKYTWGLSFAFPFYARLDN 311  
QY 1319 QGRSNARFPQNNPKWLVDFQKVMKVTVGTVTQGVKSLTSMYKVFLLISSQDGHQWT 1378  
DB 312 QGKFNWTAQTNSASEWLQIDLGSKRQVGTIITQGARDRFGHLYVAARVAVYGDGVVWT 371  
QY 1379 LFFQNG--KVYFQNGQDSFTPVNSLDPLLTLYRIHPQSWHQAIALMEVLGC 1432  
DB 372 EYNDPGASESKIFPGNNDNNHKKNIETFPQARFVRIQFVAWHNRIILRVLLGC 427  
RESULT 14  
MFCM HUMAN  
ID MFCM HUMAN STANDARD; PRT; 387 AA.  
AC Q08431;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-8) (HMFG)  
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].  
GN MFG8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI_TaxID=9606;  
RN [1] SEQUENCE FROM N.A.  
RP TISSUE=Breast, and Breast carcinoma;  
RX MEDLINE=96213908; PubMed=8639264;  
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;  
RT "Cloning and sequence analysis of human breast epithelial antigen  
BA46 reveals an RGD cell adhesion sequence presented on an epidermal  
growth factor-like domain";  
RL DNA Cell Biol. 15:281-286(1996).  
RN [2]  
RP SEQUENCE OF 170-387 FROM N.A.  
RX TISSUE=Mammary gland;  
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,  
Ceriani R.L.;  
RT "A Mr. 46,000 human milk fat globule protein that is highly expressed  
in human breast tumors contains factor VIII-like domains";  
RL Cancer Res. 51:4994-4998(1991).  
RN [3]  
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RX TISSUE=Milk;  
RA Godovac-Zimmermann J.;  
RT "Isolation and characterization of full and truncated forms of human  
breast carcinoma protein BA46 from human milk fat globule membranes";  
RL J. Protein Chem. 17:143-148(1998).  
RN [4]  
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.  
RX MEDLINE=993442076; PubMed=10411933;  
RA Haegqvist B., Naeslund J., Sletten K., Westmark G.T., Mucchiano G.,  
Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;  
RT "Medin: an integral fragment of aortic smooth muscle cell-produced  
lactadherin forms the most common human amyloid";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=97405985; PubMed=9260929;  
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;  
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein  
expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp  
(RGD)-dependent cell adhesion";  
RL DNA Cell Biol. 16:861-869(1997).  
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS  
SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.  
CC -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.  
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN  
MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.  
CC -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT  
AMINO ACID 264 AND 273.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
DR EMBL; U58516; AAC50549.1; -;  
DR EMBL; S56151; AAB19771.1; -;  
DR PIR; A47285; A47285.  
DR HSSP; P08709; 1BF95.  
DR HGNC; 7036; MFG8.  
DR MIM; 602281; -;  
DR GO; GO:0007155; P:cell adhesion; TAS.  
DR GO; GO:0007048; P:oncogenesis; TAS.  
DR InterPro; IPR001438; EGF_II.
```

```

DR InterPro: IPR006209; EGF like.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR006210; IEFG.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00754; F5_F8_type_C_2.
DR PRINTS; PR00010; EGFBL00D.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 268 317
FT DOMAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 45 48
FT SITE 27 38
FT DISULFID 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
FT CARBOHYD 350 350
SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;

Query Match 7.6%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 3.9e-31;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;

QY 1093 EMLPSKAGRWVECLGEHAGMSTLFVYSNKCOTPLGMASGHRDFOITASG---- 1147
DB 48 DYFPPS---YTCTCLAG---YAGNHC-----ETKCVPELGMENGIANSQIASSVRYTF 95

QY 1148 -QYGWAPKLRIHYSGSNAM--SYKEPFWITKVDLLAPMIHGKTQGAQKPSLLYI 1204
DB 96 LGLQHVPELRLINRAGMVTNMTPPSSNDNPMIQVNLRLRMVTVGVTCASRLASHEYL 155

QY 1205 SDFIIMYSIDGKKWQYRGNSTGLMVFFGNVDSSGKIKENIPPIIARYIIRHPTYSI 1264
DB 156 KAFKVAISLNGHEFD-FIHVNKKHKEFVGNMKNNAHVNLFPETVPAQYVRLYPTSCHT 214

QY 1265 RSTLRMLMGCDLNSCMPLGMESKAISDAITASSYF---TNMFATWSPSKARLHLQG 1320
DB 215 ACTLRPELGLGCELCGNCANPLGLKNNISIPDKQITASSSYKTGWLHLF-SNWPVSANLDKQG 273

QY 1321 RSNARPEQVNNKEMLOVDFOKTMKVTGTTQGVSKSLTSMYVKEPLISSQDGHQWTLF 1380
DB 274 NFNAWVAGSYGNDQMLQVLDGSKSEKVTGTTQGVNFGSVQVFSYKVAISNDSANWTYI 333

QY 1381 F--QNGKVKVFGQNDQSFPPVNSLDPPLLTLYRLHPSQSWHOJALRMEVLGC 1432
DB 334 QDPRIGSSKIFPGQNDWHDHSHKKNLEFETPIIARYVRLPVAWNRIRALRLELGC 387

RESULT 15
NRP2 HUMAN
ID NRP2 HUMAN STANDARD; PRT; 931 AA.
AC O60452; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2).
GN NRP2 OR VEGF165R2.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III."
RL Neuron 19:547-559 (1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RX TISSUE=Breast.
RA MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745 (1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045 (2000).
CC - FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC - SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=A22;
CC IsoId=O60462-1; Sequence=Displayed;
CC Name=A0;
CC IsoId=O60462-2; Sequence=VSP_004342;
CC Name=A17;
CC IsoId=O60462-3; Sequence=VSP_004341;
CC - SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC - SIMILARITY: Contains 2 CUB domains.
CC - SIMILARITY: Contains 2 F5/8 type C domains.
CC - SIMILARITY: Contains 1 MAM domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AF022859; AAC51788.1; -
DR EMBL; AF022860; AAC51789.1; -
DR EMBL; AF016098; AAC12922.1; -
DR HSSP; P12259; ICZT.
DR Genew; HGNC:8005; NRP2.
DR MIM; 602070; -.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; TAS.
DR GO; GO:0007411; F:axon guidance; TAS.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR00421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C_2.
DR Pfam; PF06629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.

```

DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS02022; FA58C_3; 2.
DR PROSITE; PS00060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
XW Alternative splicing.
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 149 267 CUB 1.
FT DOMAIN 277 427 F5/8 TYPE C 1.
FT DOMAIN 434 592 F5/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT DISULFID 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 809 813 Missing (in isoform A17).
FT VARSPLIC 809 830 Missing (in isoform A0).
FT CONFLICT 602 602 E -> K (in REF. 1).
FT SEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;
Query Match 6.1%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.5%; Pred. No. 8.1e-23;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;
QY 934 DEFDCK-AWAYESDVLEK-----DVHSGLIGLVCHTML-----NPAHGRQVTVQEP 983
DB 79 EXHCKYDFEIRDGSESADLLGKHCNIAPTIISGSMYIKRFTSYARQGA--GFS 136
QY 984 LEFTTFDETKSYFTFEMERNCRAPCNQIMEDPTFKENTRFHAINGYIMDTLPGLVMAQD 1043
DB 137 LRYEIP-KTGS-----EDCSKNFTSP-NGTIESPGPEKYP----- 170
QY 1044 QRIRWYLLSMGSGNETHSHFSGHVFTRKKEEYKMAVLYPGVFETVEMLPKAG--- 1100
DB 171 -----HNLDCFTILAKPKMEIILQFL---IFD-LEHDPLOVGECD 207
QY 1101 -----IW-----RVECLIGEH-----LHAGNST-----LFLV 1122
DB 208 CKYDMLDIWDGIPHVGPLIGKYCGTKTPELSASTGILSLTFTDMAVAKDGFSAFYIIV 267
QY 1123 YSN-----KCOTPLGASGHIRDFOITASQY--GQWAPKLARIHYSGSINAW-----ST 1170
DB 268 HOEPLFNFCNVPLGMSGRIANEQISASTSYSDGRWTPQOSRLH--GDDNGWTFNLDN 325
QY 1171 KEPPFSWKVLLAPMIHGIKTQGA--ROKFSLSYISQFIIMVSLDGKKQVYRGNSGTGT 1228
DB 326 KE---YLQVDFLFLMTALTAQALSRQTQNGYVYKYLEVSTNGEDWVYRHKNH- 381
QY 1229 LMVFFGNVDSGKKNHFNPIIARYIRLHPHYTHYSIRSTLRMLMGCDLNS--CSMPLGM 1286
DB 382 -KVFOANDDATEVNLKHLAPLITRFVIRPQVTHSGIALRLLELFGCRVTDAPCCSNMLGM 440
QY 1287 ESKATLSDAQITASSYFTNFWFATWSPSKAELHLQGRSNW--RPQVNNPKHLYQVDFQKT 1343
DB 441 LSLGIADSQISASS--TQEY-LWSPSAARL-VSSRS-GWFPRIPOQPFCEEWLQVLDLGT 495

QY 1344 MKVTGVTQ3-----VKSLTSMVYKBFLLISSQDGHQWTLFFQNGKV---KVFQGNQD 1394
DB 496 KTVKGVIIQGARGGDSITAVEARAFVRKFKVSYSLNGKOWE-YIQDPRTOQPKLPFGNMH 554
QY 1395 SFTPVVNSLDPELLTRYLRHHPQSMVHQ-IALRMEVLGCEAOD 1436
DB 555 YDTPDIRFD-PIPAQYVRVPERWSPAGIGRLEVLGCDWTD 596

Search completed: December 9, 2003, 17:03:10
Job time : 60 secs

Db	20	ATRYKYLGAVELSWDMQSDLLSALHADTSFSSRVPGLPLTTSVTRYKTVFVEFTDOLF	79
QY	60	NIAPRPPWMLGPTTQABVYDVTWITLKNMASHPVSLHAVGVSWKASEGAYEDDOTS	119
Db	80	NIAPRPPWMLGPTTQABVYDVTWITLKNMASHPVSLHAVGVSWKASEGAYEDDOTS	139
QY	120	QREKEDDKVPPGSHSYVVOVLKXNGPMASDPICLTYSYLSHVDLVKDLNSGLIGALLVC	179
Db	140	QKEKEDDNVTPGESHTYVVOVLKXNGPMASDPICLTYSYLSHVDLVKDLNSGLIGALLVC	199
QY	180	REGSLAKETQTLHKETLLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTVNGYVN	239
Db	200	REGSLAKETQTLHKETLLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTVNGYVN	253
QY	240	RSLEGLIGHKRSYVWHVIGMTTPVHSIFLECHTFLVNRHQASLEISPIFLTAQTL	299
Db	254	RSLEGLIGHKRSYVWHVIGMTTPVHSIFLECHTFLVNRHQASLEISPIFLTAQTL	313
QY	300	LMDLGQFLLECHLSSHOHDMAYVYKVDSCPEBPQLRMKNNBEADYDDDLTDMSEMDVVR	359
Db	314	LMDLGQFLLECHLSSHOHDMAYVYKVDSCPEBPQLRMKNNBEADYDDDLTDMSEMDVVR	372
QY	360	FDDNBSFQIRSVAKKHPTWVHYIAABEEDWDVAPLVLPDDRYSYKSOYLNNGPORI	419
Db	373	FDDNBSFQIRSVAKKHPTWVHYIAABEEDWDVAPLVLPDDRYSYKSOYLNNGPORI	432
QY	420	GRYKVKRWMAYTDEFTKTBRAIOHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG	479
Db	433	GRYKVKRWMAYTDEFTKTBRAIOHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHG	492
QY	480	ITDVRPLYSRLPKGVKHLKDFPLRGEIFKYKWTVTVEDGPKSDPRCLTRYSSFVNM	539
Db	493	ITDVRPLYSRLPKGVKHLKDFPLRGEIFKYKWTVTVEDGPKSDPRCLTRYSSFVNM	552
QY	540	ERDLASGLIGPLIICYKESVDQGNQIMSDKRVILFVSFDENRWYLTENIORELPNPA	599
Db	553	ERDLASGLIGPLIICYKESVDQGNQIMSDKRVILFVSFDENRWYLTENIORELPNPA	612
QY	600	GVQLEDPFOASIMHSINGYVFDLSQLSVCLHEVAYWYILSTGAOTDLSVFPFGYTFK	659
Db	613	GVQLEDPFOASIMHSINGYVFDLSQLSVCLHEVAYWYILSTGAOTDLSVFPFGYTFK	672
QY	660	HKMVEYEDTLPLFPFGSETVPMENPGLWILGCHNSDFRNGMTALLKYSSCDKNTGYV	719
Db	673	HKMVEYEDTLPLFPFGSETVPMENPGLWILGCHNSDFRNGMTALLKYSSCDKNTGYV	732
QY	720	EDSYEDISAYLLSKMAIEPRSP	742
Db	733	EDSYEDISAYLLSKMAIEPRSP	742
QY	743	EDSYEDISAYLLSKMAIEPRSP	742
Db	793	IKAQSVSSDILLMLGQNPTRGLFLSDUREATDRADHSRGAIERNKGPPEVASURPEL	852
QY	743	IKAQSVSSDILLMLGQNPTRGLFLSDUREATDRADHSRGAIERNKGPPEVASURPEL	852
Db	853	RHSEDEFTPEPELQRLNENLCTNITVELKKDLKXISSSDSLATSPITPSDKLAAATE	912
QY	743	RHSEDEFTPEPELQRLNENLCTNITVELKKDLKXISSSDSLATSPITPSDKLAAATE	912
Db	913	KTGSLGPPNMSVHFNHSLGTIVFCNNSSHLIQSGVPLELSEEDNDSKLEAPLNNIQESS	972
QY	743	KTGSLGPPNMSVHFNHSLGTIVFCNNSSHLIQSGVPLELSEEDNDSKLEAPLNNIQESS	972
Db	973	LRNVLMSERNLUFKEERIRGPASLKDNLKFNKVNSSVKTNRAPVNLFTNRKTRVAIPT	1032
QY	743	LRNVLMSERNLUFKEERIRGPASLKDNLKFNKVNSSVKTNRAPVNLFTNRKTRVAIPT	1032
Db	1033	LLJENSTSVQDILMERNTTEFEVTSILHNETPMDRNTTALGLNHVSNKTLTKSNVEMAH	1092
QY	743	LLJENSTSVQDILMERNTTEFEVTSILHNETPMDRNTTALGLNHVSNKTLTKSNVEMAH	1092
Db	1093	QKEDDPVPLAENPDLSKSIPLFPDWIKTHGKNLSSEQRSPKQTSIGSEKSVKQDN	1152
QY	743	QKEDDPVPLAENPDLSKSIPLFPDWIKTHGKNLSSEQRSPKQTSIGSEKSVKQDN	1152
Db	1153	FLSEKVVVEDEFTKDTLQELQIFPNKXSIFFANLANVOENDTYNOEKSPREIERKEKL	1212
QY	743	FLSEKVVVEDEFTKDTLQELQIFPNKXSIFFANLANVOENDTYNOEKSPREIERKEKL	1212
Db	1213	TQENVALPOAHMTIGTKNFKNLFLLSKQNVAGLEBQPYTPILODTRSLNDSPHSEGHK	1272
QY	743	TQENVALPOAHMTIGTKNFKNLFLLSKQNVAGLEBQPYTPILODTRSLNDSPHSEGHK	1272
Db	1273	MANFSKIREANLEGLNGQTNQVVERFPSTTRMSSNASQHVITQORGRSLKQPRLSQGBI	1332
QY	743	MANFSKIREANLEGLNGQTNQVVERFPSTTRMSSNASQHVITQORGRSLKQPRLSQGBI	1332
Db	1333	KPERKVIANDTSTQWSKNKNYLAQGLTQIEYNEKEKRAITQISPLSDCSMRNHVITQMD	1392
QY	743	KPERKVIANDTSTQWSKNKNYLAQGLTQIEYNEKEKRAITQISPLSDCSMRNHVITQMD	1392
Db	1393	SALPVAKESAPSVRHTDLTKIPSHNSHLPASACNVTFRERTSGVQBSHFLQEAEN	1452
QY	743	SALPVAKESAPSVRHTDLTKIPSHNSHLPASACNVTFRERTSGVQBSHFLQEAEN	1452
Db	1453	NLSLAPVILGITEGQOKFSSLSKSNATNPMYKULENTVLLQPELSETSKVELLSQVHVD	1512
QY	743	NLSLAPVILGITEGQOKFSSLSKSNATNPMYKULENTVLLQPELSETSKVELLSQVHVD	1512
Db	1513	QEDSPPTKTSNDSPGHLDLWGKIPLQKTQGPVPMKNTNSPGKVPFLKWAATESSEKIPSKL	1572
QY	743	QEDSPPTKTSNDSPGHLDLWGKIPLQKTQGPVPMKNTNSPGKVPFLKWAATESSEKIPSKL	1572
Db	1573	LGVLAWDNHYDTQIPSEBWKSKQSTNTAFRRKDTILPLGCENNDSTAAINEGQDKPQ	1632
QY	743	LGVLAWDNHYDTQIPSEBWKSKQSTNTAFRRKDTILPLGCENNDSTAAINEGQDKPQ	1632
Db	1633	REAMWAKQEPGRQLCSQMPVSKHOREITVTLQPEBDFKFDYDFTFSIEMKREDFDYG	1692
QY	788	EDENQSPSFQKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVPOPKVVFOEFTDG	847
Db	1693	DYENQGLRSFQKTRHYFIAAVERLWDYGMSSSHVLRNRAQSGSVPOPKVVFOEFTDG	1752
QY	848	SFTQPLXGELNEHLGLGPIYARAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDORQAE	907
Db	1753	SFTQPLXGELNEHLGLGPIYARAEVEDNIMVTFNQAASRPYSFSSLSIYSEEDORQAE	1812
QY	908	PRKNFVKNETKTYFWKQHHMAPTKDEFCKAWAYFSDVDLEKDVHSGSLIGPLVCHTN	967
Db	1813	PRKNFVKNETKTYFWKQHHMAPTKDEFCKAWAYFSDVDLEKDVHSGSLIGPLVCHTN	1872
QY	968	TUNPAHGRQVTVQEPALFTTIPDETQSWYFTENMERNCRAPCNIQMEDPTFKENTRFAI	1027
Db	1873	TUNPAHGRQVTVQEPALFTTIPDETQSWYFTENMERNCRAPCNIQMEDPTFKENTRFAI	1932
QY	1028	NGYIMDTLPLGVAQDQRIWYLLSWGNSNENIHSIFSGHVTFVKKKEYKMAVNLVPG	1087
Db	1933	NGYIMDTLPLGVAQDQRIWYLLSWGNSNENIHSIFSGHVTFVKKKEYKMAVNLVPG	1992
QY	1088	VETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMAGSHIRDOFITASG	1147
Db	1993	VETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPGLMAGSHIRDOFITASG	2052
QY	1148	QYQWAPKARLHSGSINAWSTKPEPFSWIIVDLPAMIHIGITQAGARQKPFSSLYISQF	1207
Db	2053	QYQWAPKARLHSGSINAWSTKPEPFSWIIVDLPAMIHIGITQAGARQKPFSSLYISQF	2112
QY	1208	IIMVSLDCKKOTYRGNSGTGLMVFFGNVDSGGIKHNFNPITARYIRLPHYTSIBST	1267
Db	2113	IIMVSLDCKKOTYRGNSGTGLMVFFGNVDSGGIKHNFNPITARYIRLPHYTSIBST	2172
QY	1268	LRMELMGCDLNSCSNPLGWSKAISSAQITASSYTFNMATWSPSPKARLHLCQGRSNWRP	1327
Db	2173	LRMELMGCDLNSCSNPLGWSKAISSAQITASSYTFNMATWSPSPKARLHLCQGRSNWRP	2232

```

QY 1328 QVANNPKWLQVDFOKMKVGTGVTGQVKSLLTSMYKEFLISSQDGHQHTLSPQNGKVK 1387
Db 2233 QANNPKWLQVDFOKMKVGTGVTGQVKSLLTSMYKEFLISSQDGHQHTLSPQNGKVK 2292
QY 1388 VFGNQDSFPPVNSLDPPLRLRYLRIHPQSWHQIALRMEVLGCEAQ 1435
Db 2293 VFGNQDSFPPVNSLDPPLRLRYLRIHPQSWHQIALRMEVLGCEAQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
ID O62730;
AC O62730;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Factor VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Spleen;
RA Gordy P.W.; Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF049489; AAC05384.1; -.
DR HSSP; P00451; ICFG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58 C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CRC64;

Query Match 81.0%; Score 6231; DB 6; Length 2343;
Best Local Similarity 53.3%; Pred. No. 0;
Matches 1238; Conservative 85; Mismatches 105; Indels 900; Gaps 4;

QY 1 ATRRYVLGAVELSDWYQSD-LGELFVDARFPPRPKSPFNTSVYKTLFVEFVHLF 59
Db 20 ATRKYVLGAVELSDWYQSDLLSALHADTSFSSRPVPSPLTSTYTRKTVFVEFDDLF 79
QY 60 NIAKRPMPWGLLQPTIOAEVDTVTITLKNMASHPVSLHAGVSVWKASGAEYDDQTS 119
Db 80 NIAKRPMPWGLLQPTIOAEVDTVTITLKNMASHPVSLHAGVSVWKASGAEYDDQTS 139
QY 120 QREKDDKVPFGSHYVWQVLKENGPMASDPLCLTYSVLSHVDLVKDLNSGLIGALLVC 179
Db 140 QREKEDDNPVPSHYYVWQVLKENGPMASDPLCLTYSYFSDVLDLVKDLNSGLIGALLVC 199
QY 180 REGSLAKETQTLHKLPIIFAVFDEGKSWHSETKNSLMQDRDAASARAWPKGHTVNGYVN 239
Db 200 KEGSLAKERTQTLQEFVLLFAVDFEGKSWHSETKNSLTQ-----AEQHELTHTNGYVN 253
QY 240 RSLPGLICHRKSVYVHVGNGTTFEVHSIFLEGHTFLVRNHRQASLSISPIITTAQTL 299
Db 254 RSLPGLTCHRSVYVHVGNGTTFEVHSIFLEGHTFLVRNHRQASLSISPIITTAQTF 313
QY 300 LMDLQGLLFCHISGHQDGMAYVKVDSCPEPQLRMKNNEAEYDDDLTDSMDYVR 359
Db 314 LMDLQGLLFCHIPSHQDGMAYVKVDSCPEPQLRMKNNEAEYDDDLTDSMDYVR 372
QY 360 FDDSSPFIQIRSVAKKPKTWVHYIAAEEDNDYAPLVLPADPRYSKYQVNLNNGPQRI 419
Db 373 FDDSSPFIQIRSVAKKPKTWVHYIAAEEDNDYAPSPGTPDNRSHKYNLNGPQRI 432
QY 420 GRKYKVPFMYATDFTFKTRAIQHSGLILGYGEVGDITLLIIFKNQASRPYNIYPHG 479

```

```

Db 433 GKYYKVRVAYVDETFKTRAIQHSGLILGYGEVGDITLLIIFKNQASRPYNIYPHG 492
QY 480 ITDVRPLYSRRLPKGVKHLKDPILPGEIFKFKWTVTVEDGPTKSDPRCIITRYSSPVM 539
Db 493 INYVTLHTRLPKGVKHLKDPILPGEIFKFKWTVTVEDGPTKSDPRCIITRYSSFINL 552
QY 540 ERDLASGLIGPLLIICYKESVDQGNQIMSDKRNVLFSVFDENRSLTENIQRFLNPA 599
Db 553 ERDLASGLIGPLLIICYKESVDQGNQIMSDKRNVLFSVFDENRSLTENIQRFLNPA 612
QY 600 GYQLEDPEFQASINIMSYVYVDSIOLSVCLHEVAYWYILSTGAQTFDLSPFSSYTK 659
Db 613 VYQPHDFEQLNIMHSINGYVFDNLQLSVCLHEVAYWYILSVGAQTFDLSPFSSYTK 672
QY 660 HEMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFNRNMGWTLTKLVSSCDKNGDY 719
Db 673 HEMVYEDTLTLFPFSGETVFMSENPGLMWILGCHNSDFNRNMGWTLTKLVSSCDKNGDY 732
QY 720 EDSYEDISAYLLSKNNAIEPRSP----- 742
Db 733 EPTYEDIPTLLNENNVIKPSFSQNSRHPSTKEKQKATTTTENDIEKIDLQSGERTQL 792
QY 743 ----- 742
Db 793 IYQSVSSDMLMLLQGNPTPRGLFLSDUREATDRADHSGAIBRNKGPFVEASLRPEL 852
QY 743 ----- 742
Db 853 RHSEDFTEPELQRLNENLGTNTVELKDLKISSSDSLMTSPITPSDKLAAATE 912
QY 743 ----- 742
Db 913 KTGSLGPPNMSVHFNGHLGTIVFGNNSHLIQSGVPLELSEBDNDKLEAPLMNIQESS 972
QY 743 ----- 742
Db 973 LRENVLSMESNRLFKEERIRGPASLIDKNALFKVNISSVKTNRAPVNLFTNRKTRVAIPT 1032
QY 743 ----- 742
Db 1033 LLIENTSVQDILMERNTEFEKVTSLIHNTEFMDRNTALGLNHSNKTILSKNVMAM 1092
QY 743 ----- 742
Db 1093 QKEDVPFLRAENPDLSKIPFLPDWIKTHGKNSLSEQRSPKQLTSLGSEKSVKQDN 1152
QY 743 ----- 742
Db 1153 FLSEKVVVGEDEFTKDELQEIFPNKKSIFFANLANVQENDTYNOEKKSLFEIERKEKL 1212
QY 743 ----- 742
Db 1213 TQENVALPQANTMIGTKNFKNLFLSTKQNVAGLEBQPYTPILQDTRSLNDSHSEGIH 1272
QY 743 ----- 742
Db 1273 MANFSKIREANLEGLNQTNQVVERFPSTTRMSSNASQHVITQKGEKSLQKPLSGEI 1332
QY 743 ----- 742
Db 1333 KFERKVIANDISTQWKNMYLAQGLTQIYEYKEKRAITQSPSLDCSMRNHVITIQND 1392
QY 743 ----- 742
Db 1393 SALPVAKESAPSVRHTDLTKIPSOHNSHLPASACNVTFRERTSGVQEGSHFLQAEKN 1452
QY 743 ----- 742
Db 1453 NLSLAFVTLGITTEGQKFSGLGSAKNQPMYKLENTVLLQPLGSETSKVELLSQHVVD 1512
QY 743 ----- 742

```

Db 1513 QEDSFPTKTSNDSPGHLDMGKIFLOKTQGVGNKNTSPGKVPFLKWTASESEKIPSKL 1572
QY 743 ----- 742
Db 1573 LGVLAWNDHYDTQIPSEBWSQKSKQNTAFKRDITLPGPCNNDSAAINEGQDKPQ 1632
QY 743 -----SQNPPLVLRHQREITRTTQLSDQEBIDYDDTISVEMKKEDFDYD 787
Db 1633 REAWMAKQGBPGRLCSQNPVPSKHQKEITVTTLQPEEDKFYDDTFSIEMKKEDFDYD 1692
QY 788 EDENQSPRSFOKTRHYFIAAVERLDYGHSSPHVLRNRAQSGSVQPKVVFQFTDG 847
Db 1693 DYEDQGLRSFOKTRHYFIAAVERLDYGHSSPHVLRNRAQSGSVQPKVVFQFTDG 1752
QY 848 SFTQPLRYGELNHLGLGPIYIRAEVEDNIMVFRNQAQSPYFYSLSISYEDDQCGAB 907
Db 1753 SFTQPLRYGELNHLGLGPIYIRAEVEDNIMVFRNQAQSPYFYSLSISYEDDQCGAB 1812
QY 908 PRKNFVKPNETKTYFKVQHMAPTKDEPCKAWAYFSVDLEKDVHSGLIGPLLVCHTN 967
Db 1813 PRKFNPNNETKTYFKVQHMAPTKDEPCKAWAYFSVDLEKDVHSGLIGPLLVCHTN 1872
QY 968 TLNPAKRGVTVCEALFFITRSTKSWYFTENWERNCRAPCNOMEDPTFKENYFPAI 1027
Db 1873 TLNPAKRGVTVCEALFFITRSTKSWYFTENWERNCRAPCNOMEDPTFKENYFPAI 1932
QY 1028 NGYIMDTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKBYKMAVNLVPG 1087
Db 1933 NGYVKTLPGLVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVRKKBYKMAVNLVPG 1992
QY 1088 VFTVEMLPKAGIWRVECLIGHLHAGSTLPLVTSNCKQTPGLWAGSHIRDFQITAG 1147
Db 1993 VFTVEMLPKAGIWRVECLIGHLHAGSTLPLVTSNCKQTPGLWAGSHIRDFQITAG 2052
QY 1148 QYQWAPKARLHYSGSINAWSTKEPFSMLKVDLLAPMIHGIKTGAKQKSSLYISOF 1207
Db 2053 QYQWAPKARLHYSGSINAWSTKEPFSMLKVDLLAPMIHGIKTGAKQKSSLYISOF 2112
QY 1208 IIMYSLDGKQWQYRGNTGTLMVFFGNVDSSGKHNINPPIIARIYRLHPTHYSIRST 1267
Db 2113 IIMYSLDGKQWYRGNTGTLMVFFGNVDSSGKHNINPPIIARIYRLHPTHYSIRST 2172
QY 1268 LRMELMGCDLNSCMLPGMESKASDAQITASSYFTNMENTWSPKARLHLOGRNAPRP 1327
Db 2173 LRMELMGCDLNSCMLPGMESKASDAQITASSYFTNMENTWSPKARLHLOGRNAPRP 2232
QY 1328 QVNNPKWQLQVDFQKTMKVTGTQGVKSLLTSMYVKEFLISSQDGHOWTLFFQNGKVK 1387
Db 2233 QVNNPKWQLQVDFQKTMKVTGTQGVKSLLTSMYVKEFLISSQDGHOWTLFFQNGKVK 2292
QY 1388 VEQGNQDSTFPVNSLDPALLTYLRIHQSWHQAALRMEVLGCBQ 1435
Db 2293 VEQGNQDSTFPVNSLDPALLTYLRIHQSWHQAALRMEVLGCBQ 2340

RESULT 3
Q90X47 PRELIMINARY; PRT: 2119 AA.
ID Q90X47
AC Q90X47
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SC:b22015.3 (Novel protein similar to vertebrate coagulation factor V and VIII).
DE SC:B22015.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd D.,

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AL590146; CAC94896.1; --
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 2.
DR SMART; SM00231; FA58C_2_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2119 AA; 240643 MW; DC0806EFA8761B6 CRC64;
Query Match 32.3%; Score 2486.5; DB 13; Length 2119;
Best Local Similarity 27.7%; Pred. No. 6.3e-177;
Matches 600; Conservative 262; Mismatches 491; Indels 815; Gaps 36;
QY 3 RRYIYGAVELSDWYQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLNFA 62
Db 28 RHYIAAVNIWDYTSQQ-----RTGQYKVVYREYN-EGFKQP 67
QY 63 KRPFPWGLGPTIQAEEVYDVITLKMAHSPVSLHAGVSVYKASGAEDVDDOTSORE 122
Db 68 KAHPSSGLGPTLRQGGDTIIVFRNWADHPCSLRPHGIAIGKQSGSLFYDNTSLLE 127
QY 123 KEDDKVFPGGSHYVQVVKENGPMASDPLCTYSYLSHVDLAKDLSGLIGALLVCRQG 182
Db 128 KNDVIQGEHTYQNDVTSVTPAADPPCTIYLSHFDIVRNYNTGLIGPMILCKKG 187
QY 183 SLAKETQTLH---KFILLFAVDEGKSWHSETKNSLMQDRDAASARAPKMTVNGYN 239
Db 188 TLDDSGNQ-IHPQBSVLLFGVFDENKSWIS-----TGDSPQLNVKYTINGYTN 236
QY 240 RSLGLIGCHRSVYVHVIGMTTPEVHSIFLEGHTFLVRHQRASLESPITFLTAQTL 299
Db 237 GSVPLDITCAHSKVSWHLLGMSRELPFSVFNQGVLLHDGHTSAVGLISGATISASMT 296
QY 300 LMDLQFLFLCHISSHQHDMGMEAYVYKVDSCPE--EPQLRMKNMBAEDVDDLTDSMDV 357
Db 297 GVHPGRWLVSISHKLEAGLHGLYNIRKDEBTAPKRLTIEQ----- 340
QY 358 VFDDDNPSFTQIRSVAKKHPKTVHVIYAAEEDWDVAPLVLAPDDRSYKQYLNNGQ 417
Db 341 -----KCSQEWYTYMAAEVIVDAPNMPENMDGDFRSYKQYKQSG 382
QY 418 RIGRYKVKVRMAYTDETFKTR--EAICHESGILGPLYGRVGTFLIIFKQASRPYN 474
Db 383 RIGKYKKAFTQYKDGKFKERAEKQKRELIGLGPVIRAIQIRIILKIVFKKASRPYS 442
QY 475 IYHGITDVRPLYSRRLPKGVKHLKDPFILPGISIPKYKWTVTVEGPKSDPRCLTRYYS 534
Db 443 IYHGLTIDKAAEGASVYQGGN--QYSVQPGETTYTWSVTEEDVPTSDPRCLTRYH 500
QY 535 SPVNMERDLASGLIGPLLCYKESVDQRGNIMSKRVILFVSFDENRSLVLTENIQRF 594
Db 501 SAVDAPRIASGLVGFLLCYKSLQKQVKKADQKQAMFTVDENKSWYQDENINTY 560
QY 595 LPNPAGVQLEDPEFQASNIMHSINGYVFDLSQ--LSVCLHEVAYWYILSIGAQTDLSVEF 653
Db 561 CSDPKKVKKDDPEFYKSNVMTINGVYVESQGLGFGHGEIVTVHVSQGDYIQTATF 620
QY 654 SGYTFKHQNVEDTTLTPFPFSGETVFMGMENPGWLICHNSDFNRNQTALLKVSXCDK 713
Db 621 YGHTPELKNRBDILSLFPMTGETITMNMVNIIGIWLASLNSHSDTKGMVFKDLECFR 680
QY 714 NTGDY-----YE-----DSYEDISAYL----- 731
Db 681 ---DIVIYEDGKFTAWKPTTNEIKKEPVRARPVDVDEYSOLFATLNLRTFNVK 737
QY 732 -----SK 733
Db 738 DEVEIIDTFLDQDGLLPIVEKSLGSSNENLHNATLQSFIEHGLMBSGLDKGSS 797

QY 734 NNAI-----EPRSPS 743
Db 798 NKVLNDSTOKALLETTTTPDSNRVALNNETDSIILDPPIVERKVSAPSKPMWEPESVT 857
QY 744 QN-----PPVLKHQBRIPT-----TLOSD--- 764
Db 858 MNFKTTEHINSSLERINAIYSPTETNINTMTHTDFTSITPPDGTGEMNFTLEDDTAL 917
QY 765 -----QBEI-----DVD-----DTISVEMK- 779
Db 918 LNSSESEPLQSNQNSNRATFQBEINAKDGTVDVSNVKNQIFKNVFGDGLSNSSKI 977
QY 780 --KEDFDIYD----- 787
Db 978 QVEEDFVLDSYFSEMSITMEYDVQKOTVKGESKETAQSQELSSTKXTYSGEILLES 1037
QY 788 ----- 787
Db 1038 PDITSAPNLSSSVLRNLSNESNESSNETLFMSNATFSDSTNATSSDSTATFADFNT 1097
QY 788 ----- 787
Db 1098 TFSNATFDFSNRISQMSDSSNATLSDSSNATLSDSSNATLSDSSNATLSDSSNATLSDS 1157
QY 788 ----- 787
Db 1158 SNATFSDSSNATFSDSSNATLFGVSYSSNTTILSDFSLESEMTEYILSSANDTIKSHSEVV 1217
QY 788 -----EDENQ----- 792
Db 1218 SNTSLSSESTENISLGLSNASSMKNDSESEEEVVIYNKNHSEAILTSHLDQKE 1277
QY 793 ----- 792
Db 1278 EHWGYSKHELHVKELPDHMKVYKDKSAANSNKKPKIEKKVYQVRVKKYGGYKMTYK 1337
QY 793 -----SPRSF----- 797
Db 1338 SKDYRKQPRSFSPRGFGPSVLTPRGSRPVSSSEDELTERPIVIGVPRRDENDYELVIPAQ 1397
QY 798 -----QKTRHYPTAANVER 811
Db 1398 DQEAEDFGLLDHPEEYVEYKDPKSTADVQALDATSQHLKMAAGDKNTRFYISVBE 1457
QY 812 LMDYGSSSPHVLNRNRAQSGVPQKVVQFBFTDGSFTQPLRGELNEHGLLGPYIRA 871
Db 1458 EMDYA-GYQORLDTAQRNPTVRKVVFRYLDSTFSIRDIGEMDEHLGLPLIKA 1516
QY 872 EVEDNIMVTFRNOASRPYSFSLISY-----EEDORQAEPRKVFKNETKTYFW 923
Db 1517 EVDQTVWVFRNRASRPYSLHANGVKYILKQMEGLSYDDSPYVYKQDDAVPENGFTTYW 1576
QY 924 KYQHMAPTKDBFCWAYFSDVLEKDVHSGLIGPLLVCHTNTLN--PAHGRQVTVQE 981
Db 1577 TINPKSGQNNESDCTWTYYSAVNPEDINSGLIGPLLVCKRKGTLKKPDRR-----E 1631
QY 982 FALFTIIDEKTSYFTENMERNCRAPCNIQMEDPTFKENYRFAINGYIMDTPLGLVMA 1041
Db 1632 FVLLPMTDENKSWLYEENRQIERKNRRVVM-DNFQDNLKFDAING-IIVSLKGLRWY 1689
QY 1042 ODORIRWLLSGNSNENIHSIFSGHVFTVRKBEYKMAVLYNLYGVFETVEMLPKAGI 1101
Db 1690 TNOLAKWHLNNGSPKDLHSVHFHGTQFINELKDRHQGVYPLLPFGFATLEMLPKPGL 1749
QY 1102 WRVECLIGELHAGMSTLF-----LVYSNKCQTPICMASGHRDFOI 1143
Db 1750 MLESEVGLSQRGQTILFLPDESYYKYVQVLSYLYIYAVCDHPGLLSGVQDEQI 1809
QY 1144 TASQGYQWAPKLARLHVGSSINAWST-KEPFSWIKVDLLAPMIHGIKTQCARQKFFSL 1202
Db 1810 TASDTRGQWYPLARLHNTKYNWSTSEPCQYLVQDFQRPVLSKVATQGAQFLTN 1869
QY 1203 YISQPIIYSLDGKQWYRGNSGT-----LNVFF-----GNVDSGKIKHNIFNP 1248

Db 1870 FVLNTISVSTDKKNIYYKGSDAVRKANIHKTVFFNARQTFEGNAEAYETKENIFFP 1929
QY 1249 PIATYIRLHPHYTHYRSTRLMELMGCDLNSCMLPGMESKAISDAQITASSYFTNMFA- 1307
Db 1930 PLIGRYVRLHPLHSYNFPTVRLEYTGCELDGCGSVPLGMEKGLIDDSKITASSVASWYSG 1989
QY 1308 TWSPSKARLHLOGRSNARPOVNNKEMLOWDFQKMTKVTGTTQGVKSLLSMYKPEL 1367
Db 1990 QHPWYARLNKQGTANAMQAKNDIQPMIQVELKEVKTITGTVTQAKSMGEMFVRSYI 2049
QY 1368 ISSSQGHQWTLFFQNG--KVKVFGQNDSPFTVNSLDPPLLTLYRLRHPOSWHQIAL 1425
Db 2050 LEYSEGRGEMKXTDDDDYBQKLFQGTNDNGQIKNYIYFPFISFIRIIPKQWQKSVTM 2109
QY 1426 RMEVLGCE 1433
Db 2110 RIELLGCD 2117

RESULT 4
Q43737
ID O43737 PRELIMINARY; PRT: 2224 AA.
AC O43737;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-WAR-2003 (TREMELrel. 23, Last annotation update)
DE Factor V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.
DR EMBL: Z99572; CAB16748.1; -.
DR HSBP: P12259; 1CZT.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00394; Cu-oxidase; 3.
DR Pfam: PF00754; F5 F8 type_C; 2.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
SQ SEQUENCE 2224 AA; 251673 MW; 1013108D49690EAB CRC64;

Query Match 31.0%; Score 2386.5; DB 4; Length 2224;
Best Local Similarity 26.0%; Pred. No. 2.2e-169;
Matches 592; Conservative 278; Mismatches 483; Indels 923; Gaps 35;

QY 3 RRYLGAHVLSWDMQSDLGELPVDARPPRPVPSFPFNTSV-YKTLFVFTVHLFNI 61
Db 32 RQFYVAAQGISWSVRP-----PTNSLNLSTVTSFKLIVREYEPY-FKK 75
QY 62 AKRPPNMLIGPTIQAEVTVVITLKNASHPVSLHAVGVSTWKSAGEYDDQTSQR 121
Db 76 EKPOSTISGLGPTLYAEVGDIIKVFHFNKADKPLSIHPQGIYSKLSSEGASYLDTFPA 135
QY 122 EKEDKVPFGSGSHYVVOVLKENGPMASDPLCLTYSLVSHVDLVKDLNSGLIGALLVCRE 181
Db 136 EKMDDAVAPREYIYEHYSEISDSGPTDHPDPLTHIYSHENLEDNPSGLIGLLICK 195
QY 182 GSLAKENTO-TLHK-FILLFAVDEGKSMHSETKNSLMQDRDAASAPAPKMHVNGYVN 239
Db 196 GTLTGEGTQXTQKQIVLLFAVDESKSWSQSS-----LMTVNGYVN 239
QY 240 RSLPGLIGCHRSYVHVIHGVGTTPVHSLFSGHTPLVRNHRQASLEISPTFLTAQTL 299
Db 240 GTMPDITVCAHDHISWHLGMSGPELFSHFNGVLEQNHKVSATILVYSATSTANMT 299

[illegible]

Db 261 TPQSNRMAINGVFNLPNLPNCAQKRVAMHFLGCMGNEIDVHTAFHGMQLTTRGHT 320
 QY 284 ASLEISPTTLTAQTLMDLQGLLCHLSSHQHGMVAVKVDSCPEBPQLRMKNNEA 343
 Db 321 DVANIFPAFVTAEMVPEWPGTWLISQVNSHFRDGMQALYKVKSCMAPPV----- 372
 QY 344 EDYDDDLTDSMDVVRPDDNDSFPIQIRSVAKHPTWVHYIAAEEEDWDYAPL----- 398
 Db 373 -----DLTGT-----KVRQYFIEAHEIQWYDYGPMCHDGS 401
 QY 399 -----VLAPDRSRSKSYLNGPQIRGRYKVRPMAYTDETKRAIQHES--GILGPL 452
 Db 402 TGNLREPG--SISDKFPQKSSRIGTVMKRYEAFQDETFOERKHLSEDHGLGLGV 459
 QY 453 LYBEGVGTLLIIFKQASRYNTPHGTIDVRLYRRRLPKGVKHLKOFFILPGIFK-- 510
 Db 460 IRAEVGDTIQVVFYNRASQPFMOPHGV-----FYEKDYEGTV--YNDGSSYPLVAKPF 512
 QY 511 -----YKWTYVEDGPTKSDPRCLTRYSSFVNMRDLASGLIGPLLCYKESVDQGRQV 566
 Db 513 EKVYRTVPPHAGPTAQDPACLTWMTFSAADPIRTNSGLVGLPVLVCRAGALGADGK 572
 QY 567 MSPKRNVLFSVDENRSMWLTENIQRFLEPNAGVQ-----EDPE--FQASNIMESING 619
 Db 573 GVDKERPLLTVDENKSWYNS-----ANQAAAMLDPRLLSIEDIEGFDQSNRMAING 625
 QY 620 YVFDLS--QSVCLHEVAYWILSIGACTDPLSFVFSGYFKHMYEDTLTLPFSGEIV 678
 Db 626 FLFSNLPRLDMCKGDTVAMHLLGLGTETDVHGMFQGTVQLGQMKGAAMLFPHTVMA 685
 QY 679 FMSMENPGLWLTGCHNSDFNRGWTALLKVSSCKNTGTYEDSYEDISAYLLSKNNAE 738
 Db 686 IMQPDNLGTETIYCOAGSHREAGRAINYVQC----- 718
 QY 739 PRSPQNPVVKHQRREITTLTQSDQEEIDYDITISVENKKEDEPDYEDENQSPRQ 798
 Db 719 -----PGHQ-----ATPRQRY 729
 QY 799 KTRHYFIAAVERLDYGMSS--PHVLNRQAQSGV-----POPKVVFQEP 844
 Db 730 QAARIYVMAEEVENDYCPDRSREWHQSEKDSYGFSLNKGDLRSRYKAVFREY 789
 QY 845 TDGFTOPLRGELNEHGLLGPYIRAEVDENIMVTFRNOASRPYSFYS--SLISVEEDQR 903
 Db 790 TDGFTFRPRPTGPEEHLGLTGLKEVGDIILTVVFKNNASRPYSVHAGVLESTVWP 849
 QY 904 QGABPRKVFKNETKTYFWKQHHMPTKDEPDKAWAYFSDVLEKDVHSLGLGPLAV 963
 Db 850 LAAB-----PGEVTVYQNIERSGPGFNDACSUYIYSADVPIKMYSGLVGLAI 902
 QY 964 CHNTLNPARGQVTVQBFALFTTIPDTKSNYFTFNM--ERNCRAPCNOMEDPTFKENY 1022
 Db 903 CQKGLFPHGGRSDMREFALELIFDENKSWYLEENAVTHGSDQPGSINLQDETPLESN 962
 QY 1023 RFAHNGYIMDTLPLGLVMAODRIWYLLSMGSENIHSHFSGHVTVRKKEVYKALY 1082
 Db 963 KHAINGKLVANLGLTNYQGERVAMVYMLANGQVDLHTIHFHAEFLFRNGENTRADVV 1022
 QY 1083 NLYPGVFETVEMLPKAGIWRVCLAGEHLHAQMTSLFLYNSKCTPLGMASGHIRDPO 1142
 Db 1023 DLFPGTFEVEMVASNPGLTMHCHVTDHVGAGMETLFTVFSR-----TEHLSPLT 1073
 QY 1143 ITASQYQWAPKARLHYSGSINASTKEPFWIKVDLAPMIHGIKTQGARQKFSLL 1202
 Db 1074 VITKETAQVPRDIE---BGNVAMLGWQIPK--NVEMLASVLV-----AI 1115
 QY 1203 YISQFIIMYSLDGRKQWYR 1222
 Db 1116 SVTLILVVALGGVWVYQHR 1135

RESULT 10

075180 PRELIMINARY; PRT; 1104 AA.
 AC 075180;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE KIAA0698 protein (Fragment).
 GN KIAA0698.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCEI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:169-176(1998).
 DR EMBL; AB014598; BAA31673.2; -;
 DR HSSP; P00450; 1KCW.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu_oxidase2.
 DR Pfam; PF00394; Cu-oxidase_3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 FT NON TER 1
 SQ SEQUENCE 1104 AA; 124257 MW; 617C753F766AA152 CRC64;
 Query Match 22.28; Score 1710; DB 4; Length 1104;
 Best Local Similarity 32.3%; Pred. No. 4.3e-119;
 Matches 397; Conservative 183; Mismatches 432; Indels 216; Gaps 25;
 QY 46 YKTLFVFTVHLFNIAKPRPPMGLLGPYIQAEVDTVVITLKNMASHPVSLHAYGVSY 105
 Db 19 YKTIKEYKDDSYTDBVAQPAWMLGVLGVLQAEVGVILHLKNPATRPTIHPGVFY 78
 QY 106 WKASEGAEVDQTSQREKEDDKYPPGSHYVQVQLKENGPMASDPLCLTYSLSHVDLV 165
 Db 79 EKSEGLYPGSGSLPKADDSVPFGSHIYNKTIPEGHAPTADPAFLTWIYHSHVADP 138
 QY 166 KDANSGILGALLVCREGL-----AKENTQLHKEFILLFAVDEBEGKSWH-----SETKNS 215
 Db 139 RDIATGLIGPLITCKRGALDQNSPPQRQDVHDHDFLLFSVVVDENSLWHLNENIATYCSDP 198
 QY 216 LMQPDRAASARAFKMTVNGYVNRSLPGLIGCHRSYVWYVIGMGTTPVHSLFIEGHT 275
 Db 199 ASVDKEDETTFQESNRMEAINGFVGNLPELNMCAQKRVAMHFLGCMGNEIDVHTAFHGM 258
 QY 276 PLVNRHQASLEISPTTLTAQTLMDLQGLLCHLSSHQHGMVAVKVDSCPEBPQL 335
 Db 259 LTRGHHTDVANIFPAFVTAEMVPEWPGTWLISQVNSHFRDGMQALYKVKSCMAPPV 318
 QY 336 RMKNBEAEYDDDLTDSMDVVRPDDNDSFPIQIRSVAKHPTWVHYIAAEEEDWDY 395
 Db 319 -----DLTGT-----KVRQYFIEAHEIQWY 339
 QY 396 AP-----VLAPDRSRSKSYLNGPQIRGRYKVRPMAYTDETKRAIQHES 446
 Db 340 GPMGHDSGTGNLREPG--SISDKFPQKSSRIGTVMKRYEAFQDETFOERKHLSEDH 397
 QY 447 --GILGPLLYGEVGTLLIIFKQASRYNTPHGTIDVRLYRRRLPKGVKHLKOFFIL 504
 Db 398 HLGLGVIRAEVGTDTIQVVFYNRASQPFMOPHGV-----FYEKDYEGTV--YNDGSSY 450
 QY 505 PGEIFK-----YKWTYVEDGPTKSDPRCLTRYSSFVNMRDLASGLIGPLLCYKES 558
 Db 451 EPLVAKFEKVTYRTVPPHAGPTAQDPACLTWMTFSAADPIRTNSGLVGLPVLVCRAGA 510
 QY 559 VDQRGNQIMSKRNVLFSVDENRSMWLTENIQRFLEPNAGVQ-----EDPE--FQAS 611

Db 511 LGADGKQGVKXFEFFLFTVLNENKSYN-----ANQAAWLPRLLEDIEGFQDS 563
 Qy 612 NIMHSINGVYVDSL-QLSVCLHEVAVYVYLSIGAQDTFLSVFFSGYTFKHVMYEDTLTL 670
 Db 564 NMHAINGFLSNLPRDMCKGDTVAWHLEGLGTETDVHGVMPQGNVQLQGRKGAAML 623
 Qy 671 PPSGSETVFMSPENPGLWILGCHNSDPNRGWTALLKVSDDCNTGTYEDSIEDISAVL 730
 Db 624 PPHTFVMAIQPDMGLGTFBYICQAGSHRAGRAIYNVSC-----564
 Qy 731 LSKNNAIEPRFSQNPVPLKRGHQBTRTTLQSDQBEIDYDDTISVEMKKEDDIYDEDE 790
 Db 665 -----PCHQ-----668
 Qy 791 NQSPRSFQKTRHYFIAAVERLWDYGMSS- PHLNRRAQSGSV-----PQR 836
 Db 669 -ATPRQYQARIYINAEVENDYCPDRSWERENHNSQSEKSYGIFLNSKDGGLGSR 727
 Qy 837 KKVYFQFTDGSFTQPLRYGELNHLGLLPYIRAEVEDNIMVTFRNOASRPYSFY-SL 895
 Db 728 KKAUFREVTDTGTRIPRPTGPEHLGILGLIKGEVGOILTUVFKNNASRPYSVHAGV 787
 Qy 896 ISYEDQORQNEPRKFNKNETKTYFKVQHMAPTKDFDCKANAYESDVLEKDVHS 955
 Db 788 LESTVWPLAAE-----PGEVVTYQWNIPEPSGGPNDPNSACVSIYYSANDPIKDMYS 840
 Qy 956 GLIGFLVLCVHTNTLNPAHGRQVTVQEPALFTTFIDETKSWYFTENN-ERNCRAPCNIOME 1014
 Db 841 GLVGLAICQKILEPHGRSDMDREFALLFLIFDENKSWYLEENAVATHSGQDPGSLNQ 900
 Qy 1015 DPTKENVYRPHANGYIMDTLPLGVMAQODIRIYLLSGNSNENIHSIPESHGVFVRKK 1074
 Db 901 DETFLESNKHAINKGLYANLRLTYQGERVAVYMLAMQDDVLEHTIHPAESFYLRNG 960
 Qy 1075 BEYKALYNLYPGVFETVEMLPKXAGIWRVECLIGERHLHAGMSTFLVYSNKKQTPLEMA 1134
 Db 961 ENYEADVVDLPFGTFEVVEMVAVSNPGTWLMCHCTVDHVGAMETLTVFSR-----1011
 Qy 1135 SGHJRDQITASQYQGWAPKLARLYHSGSINAWSTPEPFSWIKVLLAPMIHGIKTQG 1194
 Db 1012 TEHLSPLUTVIITKETEKAAPPDRIE---EGNVKMLGMQIPIK--NVEMLASVLV-----1059
 Qy 1195 ARQKFSLSYISQFIIMYSLDGKKWQTYR 1222
 Db 1060 -----ALSVTLVLLVVALGCVVYQHR 1081

RESULT 11

Q9JL97 ID Q9JL97 PRELIMINARY; PRT; 1084 AA.
 AC Q9JL97;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE GPI-anchored ceruloplasmin.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 EN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=Sprague Dawley; TISSUE=Brain;
 EX MEDLINE=20127919; PubMed=10660599;
 RA Patel B.N., Dunn R.J., David S.;
 RT "Alternative RNA Splicing Generates a Glycosylphosphatidylinositol-
 anchored Form of Ceruloplasmin in Mammalian Brain";
 EL J. Biol. Chem. 275:4305-4310(2000).
 DR EMBL; AF202115; AAF34175.1; --
 DR HSSP; P00450; IKCW.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu oxidase.
 DR Pfam; PF00394; Cu-oxidase; 3.

DR PROSITE: PS00079; MULTICOPPER OXIDASE1; 3.
 DR PROSITE: PS00080; MULTICOPPER OXIDASE2; 1.
 SQ SEQUENCE 1084 AA; 123749 MW; 956584154257C55 CRC64;
 Query Match 22.2%; Score 1709.5; DB 11; Length 1084;
 Best Local Similarity 32.7%; Pred. No. 4.6e-119;
 Matches 384; Conservative 188; Mismatches 417; Indels 185; Gaps 22;
 Qy 3 RRYVLGAVELSDWYMQ-SDLGEL-PVDARPPRPVKSPPEPNTSVVYKKTILFVRETVHLEN 60
 Db 22 KHYTIGITEAVNDVASGEKEKELISVDETSNPNFLRMGPDRIGRKYKALYSEITDGTFT 81
 Qy 61 IAKPRPPMGLGFTIOAEVYDVTYVITLKNMASHPVSLHAVGVSYWKASGAEYDDQTSQ 120
 Db 82 KTIIDPAWLGILGVPINAEVGVSVHVGNFASRPYTFHAHGVYTYTKANEGAIYPDNTD 141
 Qy 121 REKEDDKVFPGGSHYVNOVLKENGPMASDPLCLTYSLSHVDLVKDLNGLIGALLVCR 180
 Db 142 FQRAADKLFQQQQLYVYLR-NEPSPGEGSDNCVTRYHSHVDAPKDIASGLIGPLILCK 200
 Qy 181 EGSIAKEKIQTL-HKFIILFAVPDGGKSMHSETKNSLM-----QDRDAASARAWPKHT 233
 Db 201 KGSIAKEKEENIDQEFVLMFESVDENLSWLEDNIKTFCEPEKVDKDNEDFQESNEMYS 260
 Qy 234 VNGYVNRSLPGLIGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISITP 293
 Db 261 INGYTFGLSLGSCAEDRVKVKYLFMGNEVDVHSELFGQALTSKNYHTDIINLFPATL 320
 Qy 294 LTAQTLMDLQQLFLCHISSHQHGMAYVYKVDSCPEEPQLRMKNNEEAEDYDDLDTS 353
 Db 321 IDVSNVAQPGVWMLSCQNLNHLKAGLQAFQVDC-----NKPSP-DDDIQDR 368
 Qy 354 EMDVVREFDDNSPSFIQIRSVAKKPKTWHYIAAEEDMDYAP-----LVLP 402
 Db 369 HV-----RH-----YYIAAEETINDYAPSGDTFTTGENLTSLGS 402
 Qy 403 DRSYKSOYLNNGPQIRGRKYKVRMAYTDETE---KTREAIQHESGILGLPILAYGVGD 459
 Db 403 DSRVFFEQ---GATRIQGSYKLVRYBYTDDSTNRKQRPDDEHGLIGLPVIAEVD 458
 Qy 460 TLLIIFKNQASRPNIYPHGITDVR---PLYSRRLPGVKVHLKDPILPGEIPKYKWTY 515
 Db 459 IIRVTFNKQGFPLSIQPMQVRFTKENEGTYG---PDGRSSKQASHVAPKETTYEWTY 515
 Qy 516 TVEDGPTKSDPRCLTRYSSFWNMEDLASGLIGLALICYKESVDQGNQIMSDKRVIL 575
 Db 516 PREMGFTYADPVCLSKMIYSGVDLTKDIFTLIGMKICKRGSLLADGRQDVNKEYL 575
 Qy 576 FSVFDENRSWYLTENIQRFLENPAGVQLDEPFOASNIMHSINGYVFDLSQ-LSVCLHEV 634
 Db 576 ATVFDENESLLDDNIRMFTTAPENVDKEDDFQESKNKMSMNGMYGNLPLGNMCLGES 635
 Qy 635 AYWTLSIGAQDTFLSVFFSGYTFKHVMYEDTLTLFPFSGEYTFVFMSPENPGLMILCHN 694
 Db 636 IVWYLFSGNADVHGIYFSGNTYLSKGRDRTANLPHKSLTLLMTPDTEGSDVECLT 695
 Qy 695 SDFRNRGHTALLKYSSCDKNTGYDYESYEDISAYLSKKNNAIEPRFSQNPVPLKRGH 754
 Db 696 TDHYTGKMKQYVYVQ-----KQPEDVTLY-----722
 Qy 755 EITRTTLOSQOEIDYDDTISVEMKKEDFDIYDEENQSPRSFQKTRHYFIAAVERLWD 814
 Db 723 -----QGETYYIAAVERVWD 738
 Qy 815 YMGSSPHVLNRRAQSGSV-----POFKKVVQEFDTGSGFTQPLVYRGELNEH 861
 Db 739 YSPREDWEMELHLQEQVNSNAFLDKSEFFIGSKYKVVYTRFTDSTFRQVRRAREEH 798
 Qy 862 LGLIGPYIRAEVEDNIMVTFRNOASRPYSFYSLISYEEDQORQAEPRKNFVK---PNET 918
 Db 799 LGILGPLIHADVGVKVKVAFKNNASRPYSIHA-----HGVTKTSVTPTLSEV 848
 Qy 919 KTYFWKVQHMAPTKDFDCKAWAYFSDVLEKDVHSLIGLPLIVC---HTNTLNPAHGR 975

```

Db 849 RTYLWQIPERSGAGTDSPCIPWAYYSTVDVVKDLYSLGLIGPLVCKRSYVYKFNPK--- 905
QY QVTQVEFALFTIIDEYKSTFTENMERNCRAPCNIQMEDPTFKENYRPHAINGYIMDL 1035
Db 906 --KRMEXSLLEFVDENESWYLDNDNINYSDHPEKVNKNDEBFESGMOHAINGKPGNL 963
QY 1036 PGLVMAQDQRTWYLLSGNSNENIHSIFSGHVFTRVKKKEVKALYNLYPGVFVEML 1095
Db 964 QGLTWGVDEVNMYVMANGNEIDLKTVFHHGHSYQYKRGHSSDVFDFLFFGYQTLMP 1023
QY 1096 PSKAGIWRVECLIGELHAGHAGSTLFLVYSNKKQT 1129
Db 1024 PQTPGTWLLHCHVTDHAGMVTYTYTLPNQASS 1057

RESULT 12
Q9XT27 PRELIMINARY; PRT; 1048 AA.
AC Q9XT27;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Ceruloplasmin.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
EX MEDLINE=99384006; PubMed=10452945;
RA Lockhart P.J., Mercer J.F.B.;
RT "Cloning and expression analysis of the sheep ceruloplasmin cDNA.";
RL Gene 236:251-257(1999).
DR EMBL; AF134814; AAD41477.1; -.
DR HSSP; P00450; 1KCW.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu oxidase.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1048 AA; 119125 MW; 925F16D07B0549CBB CRC64;

Query Match 21.0%; Score 1618; DB 6; Length 1048;
Best local similarity 32.2%; Pred. No. 3.2e-112;
Matches 378; Conservative 177; Mismatches 417; Indels 202; Gaps 25;

QY 3 RRYLGAVELSDWYQSDLGE---LPVDARPPRPVPSFPPTSVYKTLFVEFTVHLF 59
Db 22 KHYTIGITETAWNY-ASDHAEEKLISVDTEHSNIVLQNGPNRIGSVYKAVLYQYDENF 80
QY 60 NIAREPWPWGLLPTIQAENVYDVTVVILKNASHPVSLHAGVSYWKASGAEAYDDQTS 119
Db 81 RTVIEKPWLGFLGPIIKAEKTDKVVYVHLKNFASRPYTFHAHGLTYKHEGALYDNTT 140
QY 120 QREKDDKVPFGSGSYTHVQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLIGALLVC 179
Db 141 DLQKADDDKVPQEQCLYLHNPENQGPGEEDSNCTRIYHSHIDAPKDIAAGLIGPLHIC 200
QY 180 REGSLAKEKTOPLHK-FILFAVDFEGKSMH-----SETKNSLMQDRDAASAWP 229
Db 201 KKDLSDEEKRIDKEFVVMFVSVVDENLSWYLEENIKTYCSEPEKVEQDNEDFQESN--- 257
QY 230 KMTVNGYNNRSLPGLIGCHRSYVTHVIGTTPTEVHSIFLEGHTFLVNRHROASLEIS 289
Db 258 RMYSVNGYAFGLSLPGLSMCAEDRVKWLFGMGNEIDVHAAPFHGQVLTSTKRYRDTINLF 317
QY 290 PITFLTAQTLMLDLGQFLFLFCHISSHQHDGAEVYKVDSCPEPQLRMKNNEAEDYDD 349
Db 318 PATLFDAMVAQNPQWMLSCQVNLHKLKAGLQAFWVQDCKSS----- 361

```

```

QY 350 LTDSEMDVVRFDNDSPSFIQIRSVAKHPKTKTWVHYIAAEEEDWDYAPL-----VL 400
Db 362 -----SEDN-----IHGKNVRH-----YVIAAEEVINTWAPSIDAFTKENLR 399
QY 401 APDRSVKSYQLANGPORIGRKYKRVFMAYTDETR---KTBRAIQHESGILGPLYGEV 457
Db 400 APGSAS--EAFPEGQFTRIGSYKLVYREYTDASFSNQKBERGPEBEHILGILGFVIAAEV 457
QY 458 GDTLLIIFKNOASFPYNIYPHGI-----TQVRLYRRLPKGVKHLKDFPILGEIF 509
Db 458 GDTIRVFNKAAHPLSIEPIGVAVDKVNEGTYVSPGSGPPSG-SH-----VAPKGT 511
QY 510 KYKTVTVVEQPTKSDPRCLTRYSSPVRMERDLASGLIGPLLCYKESVDQGNQMSD 569
Db 512 TYBWTVPKEVGPTYKDPVCLAKMYIS--GSTKDIFTGLIGPMKICRNGSLANGRLKRVND 569
QY 570 KRNVLFSVFDENRSWYLTENIQFLPNPAGVQLEPOEFQASNIHMSIINGYVDSLSQ-LS 628
Db 570 KEFVLPFTVFDENESLLDDNIMFTTAPQVDKENEDEFSKNKMSKMGFMIGNQGLS 629
QY 629 VCLHEVAWYILSIGAOTDFLSVFFSGYTFKHQVYEDTLTLPFSGEIVFENSMENPGLM 688
Db 630 MCQGSVMWYILFSAGNEVDIHGYISGNTYLSGERDRTANLPQTSLSLFMQPOTAGTF 689
QY 689 ILGCHNSDFRNRGMTALLKYSSCDKNTGDYVEDSYEDISAYLLSKNNAIEPRFSQNPV 748
Db 690 DVSCLTTHYTGKWKQKTYVSCQGRS-----EDLYLYL----- 723
QY 749 LKRHQREITRTTLOSQDEBIDYDITISVEMKEDFDIYDEDENQSPRSFQKTRHYFLAA 808
Db 724 -----GERTYIIAA 732
QY 809 VERLWDYGMS---SSPHVLNRRAQSGSV-----POFKKVVFOEFTDGSFTOPLVR 855
Db 733 VEVEDMSPSKWEKELHLQESQNLNAPLXDEEFYIGSYKAVVVRQFDTSTFQVPPER 792
QY 856 CELNEHLGLGPIYIRAEVEDNIMVTFRNQASRPVSFYSSLSISVEEDQOQASPRKNFVKP 915
Db 793 XGEBEHLGILGQLHADVGOKVNIIFKNMATRPSYIHAHGVKTESST---VTP---TAP 845
QY 916 NETKTYFWKQVHMAPTKDEDFCKAWAYFSDVLEKDVHSGLIGPLLVCHTNLT---NPA 972
Db 846 GETRIYIKIPERSGAGWDSPCIPWYYSYTVDEKDLFSGLIGPLIVCKHLYKSNP- 904
QY 973 HGRQVTVQEPALFTTIFDETKSWYFTTENMERNCRAPCNIQMEDPTFKENYRPHAINGYIM 1032
Db 905 ----IKLEFSLLEFVDENESWYLDNDNIKTYSHPKVKDKANEERFESKHAINGRMF 960
QY 1033 DTLPLGVMAQDQRTWYLLSGNSNENIHSIFSGHVFTRVKKKEVKALYNLYPGVFETV 1092
Db 961 GNLQGLTW-----HVGNEVDLHRSVHRHSHFSQYQHRGLIYSDVDFLFFGYQIL 1009
QY 1093 EMLPSKAGIWRVECLIGELHAGHAGSTLFLVYSNKK 1126
Db 1010 EMTPKTPIGILLHCHVTDHAGMVTYTYTLPNE 1043

RESULT 13
Q8BQ43 PRELIMINARY; PRT; 355 AA.
AC Q8BQ43;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Coagulation factor VIII (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
MEDLINE=22354683; PubMed=12466851;

```

Wed Dec 10 11:56:49 2003

RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK051579; BAC34681.1; -.
 FT NON TER 355
 SQ SEQUENCE 355 AA; 40548 MW; 2D3C984EFA43284 CRC64;

Query Match 20.0%; Score 1534.5; DB 11; Length 355;
 Best Local Similarity 85.4%; Pred. No. 1e-106;
 Matches 286; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

QY 1 ATRRYLGAVALSMDYMSD-LGELPVDARPPPPVPPVKSFPPTSVVVKTLFVETVHLF 59
 DB 20 ATRRYLGAVALSMDYMSD-LGELPVDARPPPPVPPVKSFPPTSVVVKTLFVETVHLF 79

QY 60 NTAKEPRPMMGLGPTIOAEVVDVITLKNMASHPVSLHAGVSVYKASEGAEYDDQTS 119
 DB 80 NTAKEPRPMMGLGPTIOWEVDVITLKNMASHPVSLHAGVSVYKASEGAEYDDQTS 139

QY 120 QREKEDKVFPGSGSHYVQVLKNGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVC 179
 DB 140 QMEKEDKVFPGSGSHYVQVLKNGPMASDPLCLITYSLSHVDLVKDLNGLIGALLVC 199

QY 180 REGSLAKETOPHLKFIILFAVDFEGKSMHSETKNSLMQDRDAASAPAWKMTVNGYVN 239
 DB 200 KEGSLAKETOPHLKFIILFAVDFEGKSMHSETKNSLMQDRDAASAPAWKMTVNGYVN 259

QY 240 RSLPLGLICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISFIFLTAQTL 299
 DB 260 RSLPLGLICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISFIFLTAQTL 319

QY 300 LMDLQGLFCHISSHQDMEAYVYKVDSCPEEQ 334
 DB 320 LMDLQGLFCHISSHQDMEAYVYKVDSCPEEQ 354

RESULT 14
 Q8C4S2 PRELIMINARY; PRT; 847 AA.

AC Q8C4S2;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE HEPHRAESTIN homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RC MEDLINE=22354683; PubMed=12466951;
 RA The FANTOM Consortium,
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK081330; BAC38197.1; -.
 SQ SEQUENCE 847 AA; 95136 MW; 778CEC0617243402 CRC64;

Query Match 16.4%; Score 1265; DB 11; Length 847;
 Best Local Similarity 31.8%; Pred. No. 7.2e-86;
 Matches 304; Conservative 129; Mismatches 308; Indels 214; Gaps 20;

QY 1 ATRRYLGAVALSMDYMSD-LGELPVDARPPPPVPPVKSFPPTSVVVKTLFVETV 56
 DB 24 ATRRYLGAVALSMDYMSD-LGELPVDARPPPPVPPVKSFPPTSVVVKTLFVETV 83

QY 57 HLFNIAKPRPPMGLGPTIOAEVVDVITLKNMASHPVSLHAGVSVYKASEGAEYDD 116
 DB 84 GTTTEBIAPWGLGFLPQLQAEVGVGVILHLKNFASFPYTHPHGVYKESGSLYD 143

QY 117 QTSOREKEDDKVFPGGSHYVQVLKNGPMASDPLCLITYSLSHVDLVKDLNGLIGAL 176
 DB 144 GSSGVLKADSDVFPGGSHYVQVLKNGPMASDPLCLITYSLSHVDLVKDLNGLIGAL 203

QY 177 LVCREGSL---AXEKTQTLHKFILLFAVDFEGKSMHSE-----TKNSLMQDRDAASAR 226
 DB 204 ITCRGLTLDGNSPPQRKXVDHNFLLSFVIDENSLMHLDDNLTATYCSDPASVDKDEGACQ 263

QY 227 AMPXMHYVNGVYNSRSLPGLICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASL 286
 DB 264 DSNEMHAIINGFVFGNLPSELSCAQKQVAMHLFCMGNEIDVHTAFPHQMLSRGHTOVA 323

QY 287 EISPIITLTAQTLMDLQGLFCHISSHQDMEAYVYKVDSCPEEQLEMKANEAEYD 346
 DB 324 NIFPATVTAEMVPPKSGTGLISCEVNSHLRSGQAFYKVDSCMDPPV----- 372

QY 347 DDDLTDSGEMVVRDNDNSPFIQIRSAVKKHFKTWHYIAAEEDWDYAPLVLPDDR 406
 DB 373 -DQLTG-----KVRQYFIOAHEIQMDYGP--IGYDGR 402

QY 407 YKS-----QVLNNGPQRIKRYKVPFMAVDTETFKTRAIQHS---GILGPLL 455
 DB 403 GKSLRPGSGDPKTFQKSSRIGTYMKVRYEAFQDETFOERVHQBEETHLGLSPVIRA 462

QY 456 EVGDTLILIFKNQASRPYNIYPHGITDVRPLYSRLLPKGVKHLKDFPLFGEIFKYN 515
 DB 463 EVGDTIQVFNASQPFSTQPHGV-----FYKCN-----SEGTYV----- 498

QY 516 TVBDGPTKSDPRCLTRYYSFVNMERDLASGLIGPLLI CYKESVDQNGNIMSDKNVIL 575
 DB 499 -NDPDT-----RDTNSGLVGLPLVCKAGALGADGKQGVKDEFFLL 538

QY 576 FSVFDENRSWLTENIQRELPNAG-----VQLEDRE-FOASMIHMSINGYVDSL--QLS 628
 DB 539 FIVFDENESWYNNAN-----QAAGMLDSRLLEDVEGFDSDNRMAINGELFENLPRLD 592

QY 629 VCLHEVAYWYILSIGAOTDPLSFVPSGYTFPKHMYEDTITLTPFSGSETVFNSEMPGLM 688
 DB 593 MCKGUTVAMHLGLGTETDVGVMFEGNTVQLQMRKGVMLPHTFTVTAIMPDPNGIF 652

QY 689 ILGCHNSDFRNRGMTALLKYSSCDKNTGDYVDSYEDISAVLLSKNAISPRFSQNPV 748
 DB 653 ELYCOAGSHREGMQAIYNSQC-----SHQD----- 680

QY 749 LKHOREITRITLQSDOEIDYDDTISVEMKKEDFDYDEENQSPRSFOKKTTHFI 808
 DB 681 -----SPRQHYQASRVYIMA 696

QY 809 VERLWDY-----GMSSSPHVLNRNRAQSGVFPQKVVFOEFTDGSFTOPLY 854
 DB 697 EIEIWDYCPDRSWELEWNTSEKDSYGHVFLSNKDGLLGSKYKKAVERVTDGTRIPRP 756

QY 855 RGLMNEHLGLGPIYRAFEVDNIMVTRNQAASRPYSYS-SLISYEDQDQAGRP 908
 DB 757 RSGPREHLGLGPIYRAFEVDNIMVTRNQAASRPYSYS-SLISYEDQDQAGRP 811

RESULT 15
 O75659 PRELIMINARY; PRT; 782 AA.

AC O75659;
 DT 01-NOV-1998 (TREMELrel. 08, Created)
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE DJ46618.1 (Coagulation factor V (Activated protein C cofactor),
 DE coagulation factor VIII (Procoagulant component) and ceruloplasmin
 DE (EC 1.4.6.3.1, ferroxidase like) (Fragment)).
 GN DJ46618.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

